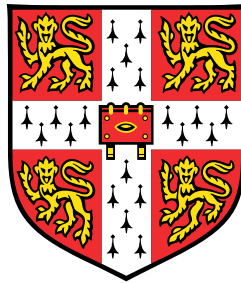


# Single atom imaging with time-resolved electron microscopy



**Thomas Furnival**

Department of Materials Science & Metallurgy  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*



## Declaration

This dissertation is submitted to the University of Cambridge in partial fulfilment of the degree of Doctor of Philosophy. It is an account of the research I have undertaken in the Department of Materials Science and Metallurgy at the University of Cambridge between October 2013 and February 2017, under the supervision of Prof. P. A. Midgley and Dr. P. D. Bristowe. The work described is original and a result of my own work, and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This dissertation does not exceed the word limit set by the Degree Committee for the Faculty of Physics and Chemistry, and no part of it has already been, or is currently being, submitted for any other degree, diploma, or other qualification, at any other university.

Aspects of this thesis have been published or submitted for publication:

- T. Furnival, R. K. Leary and P. A. Midgley, “Denoising time-resolved microscopy sequences with singular value thresholding”, *Ultramicroscopy*, vol. 178, pp. 112–124, 2017.
- Z. Chen, S. Mitchell, E. Vorobyeva, R. K. Leary, R. Hauert, T. Furnival, Q. M. Ramasse, J. M. Thomas, P. A. Midgley, D. Dontsova, M. Antonietti, S. Pogodin, N. López and J. Pérez-Ramírez, “Stabilization of Single Metal Atoms on Graphitic Carbon Nitride”, *Adv. Func. Mater.*, vol. 27, no. 8, p. 1605785, 2017.
- T. Furnival, R. K. Leary, E. C. Tyo, S. Vajda, Q. M. Ramasse, J. M. Thomas, P. D. Bristowe, P. A. Midgley, “Anomalous Diffusion of Single Metal Atoms on a Graphene Oxide Support”, *Chem. Phys. Lett.*, vol. 683, pp. 370–374, 2017.
- T. Furnival, D. Knez, E. Schmidt, R. K. Leary, F. Hofer, P. D. Bristowe, P. A. Midgley, “Adatom dynamics and the surface reconstruction of Si(110) revealed using time-resolved electron microscopy”, *Manuscript submitted*.

Thomas Furnival  
December 2017





## Acknowledgements

First and foremost, I would like to thank my supervisors Paul Midgley and Paul Bristowe for their guidance, advice and enthusiasm throughout the course of my studies. Above all, the freedom and encouragement to explore seemingly disparate topics has helped to formulate the research presented herein. This work would also not have been possible without the assistance of Rowan Leary, who has patiently guided me through the minefield that is aberration-corrected electron microscopy. Many of the figures, videos and discoveries in this work would not have been possible without his efforts on the Titan.

I must thank the members of the Electron Microscopy Group here in Cambridge. Duncan Johnstone and Ben Martineau have been particularly helpful in discussions about all aspects of machine learning. Sir John Meurig Thomas has opened my eyes to the important role played by electron microscopy in the field of catalysis. Elsewhere in the department, thanks go to Eric Schmidt for the work on particle detection and tracking, and to Georg Schusteritsch and Chris Pickard for their help with DFT and AIRSS simulations. Eric Tyo and Stefan Vajda from Argonne National Laboratory provided the copper cluster samples that inspired much of the work in this thesis. Quentin Ramasse at SuperSTEM acquired some of the high-resolution STEM data for this thesis, and thanks also to Daniel Knez and Ferdinand Hofer from Graz for the silicon datasets that extended the time-resolved work to crystalline systems.

Finally, I would like to thank all of those who have kept me motivated me along the way, and special mentions go to Steve, Andy, Rob, James and Robin. Most importantly of all, to Dad, Josh and Jen – your endless support has been a source of inspiration for which I will be ever grateful.



## Abstract

Developments in scanning transmission electron microscopy (STEM) have opened up new possibilities for time-resolved imaging at the atomic scale. However, rapid imaging of single atom dynamics brings with it a new set of challenges, particularly regarding noise and the interaction between the electron beam and the specimen. This thesis develops a set of analytical tools for capturing atomic motion and analyzing the dynamic behaviour of materials at the atomic scale.

Machine learning is increasingly playing an important role in the analysis of electron microscopy data. In this light, new unsupervised learning tools are developed here for noise removal under low-dose imaging conditions and for identifying the motion of surface atoms. The scope for real-time processing and analysis is also explored, which is of rising importance as electron microscopy datasets grow in size and complexity.

These advances in image processing and analysis are combined with computational modelling to uncover new chemical and physical insights into the motion of atoms adsorbed onto surfaces. Of particular interest are systems for heterogeneous catalysis, where the catalytic activity can depend intimately on the atomic environment. The study of Cu atoms on a graphene oxide support reveals that the atoms undergo anomalous diffusion as a result of spatial and energetic disorder present in the substrate. The investigation is extended to examine the structure and stability of small Cu clusters on graphene oxide, with atomistic modelling used to understand the significant role played by the substrate. Finally, the analytical methods are used to study the surface reconstruction of silicon alongside the electron beam-induced motion of adatoms on the surface.

Taken together, these studies demonstrate the materials insights that can be obtained with time-resolved STEM imaging, and highlight the importance of combining state-of-the-art imaging with computational analysis and atomistic modelling to quantitatively characterize the behaviour of materials with atomic resolution.



# Contents

<b>Nomenclature</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Single atom electron microscopy</b>	<b>5</b>
2.1 Introduction . . . . .	5
2.2 Scanning transmission electron microscopy . . . . .	6
2.3 Alternative methods for single atom imaging . . . . .	13
2.4 What are the challenges for single atom STEM? . . . . .	15
2.5 Summary . . . . .	19
<b>3 Low-rank methods for electron microscopy</b>	<b>21</b>
3.1 Introduction . . . . .	21
3.2 Matrix factorizations in electron microscopy . . . . .	22
3.3 Singular value thresholding for denoising . . . . .	28
3.4 Robust principal component analysis . . . . .	40
3.5 Summary . . . . .	48
<b>4 Tracking the motion of single atoms with ADF-STEM</b>	<b>51</b>
4.1 Introduction . . . . .	51
4.2 Robust detection of single atoms . . . . .	52
4.3 Localizing and characterizing single atoms . . . . .	58
4.4 Automated tracking of atomic motion . . . . .	65
4.5 Summary . . . . .	72
<b>5 Anomalous diffusion of copper atoms on graphene oxide</b>	<b>73</b>
5.1 Introduction . . . . .	73
5.2 Atomic structure of graphene oxide . . . . .	74
5.3 Capturing single atom dynamics with ADF-STEM . . . . .	79
5.4 Analyzing and modelling atomic diffusion . . . . .	82
5.5 A hybrid model for adatom diffusion on GO . . . . .	86
5.6 Summary . . . . .	97

<b>6</b>	<b>Structures of small copper clusters on graphene oxide</b>	<b>99</b>
6.1	Introduction . . . . .	99
6.2	ADF-STEM results . . . . .	100
6.3	Distribution of copper adatoms on GO . . . . .	104
6.4	Computational structure searching . . . . .	110
6.5	Structures of small copper clusters . . . . .	112
6.6	Summary . . . . .	123
<b>7</b>	<b>Adatom dynamics and the Si(110) surface</b>	<b>125</b>
7.1	Introduction . . . . .	125
7.2	Silicon surface studies . . . . .	126
7.3	ADF-STEM results . . . . .	128
7.4	Imaging the silicon (110) surface . . . . .	129
7.5	Pt adatom diffusion on Si(110) . . . . .	136
7.6	Summary . . . . .	138
<b>8</b>	<b>Conclusions and future work</b>	<b>139</b>
	<b>Appendix A Supplementary Materials</b>	<b>145</b>
A.1	Supplementary Movies . . . . .	145
A.2	Supplementary Software . . . . .	147
	<b>Appendix B Poisson–Gaussian unbiased risk estimator</b>	<b>149</b>
B.1	Stein’s Lemma . . . . .	149
B.2	Hwang’s Lemma . . . . .	150
B.3	Poisson–Gaussian unbiased risk estimator . . . . .	150
	<b>Appendix C Stochastic gradient descent</b>	<b>153</b>
C.1	Theory . . . . .	153
C.2	Derivation of the update step . . . . .	154
C.3	Choosing the step size . . . . .	154
C.4	Accelerating stochastic gradient descent . . . . .	155
	<b>Appendix D Density functional theory</b>	<b>157</b>
D.1	Theory . . . . .	157
D.2	Calculation details . . . . .	158
	<b>References</b>	<b>161</b>

# Nomenclature

## Abbreviations

ABF	(Annular) Bright-Field
ADF	Annular Dark-Field
AIC	Akaike Information Criterion
AIRSS	<i>Ab initio</i> Random Structure Searching
BIC	Bayesian Information Criterion
CRLB	Cramér-Rao Lower Bound
CSR	Complete Spatial Randomness
CTRW	Continuous-Time Random Walk
DFT	Density Functional Theory
EDS	Energy-Dispersive X-ray Spectroscopy
EELS	Electron Energy Loss Spectroscopy
FIM	Field Ion Microscopy
GGA	Generalized Gradient Approximation
GO	Graphene Oxide
LDA	Local Density Approximation
LoG	Laplacian-of-Gaussian
(K)MC	(Kinetic) Monte Carlo
MD	Molecular Dynamics

---

MLE	Maximum Likelihood Estimation
MSD	Mean-Squared Displacement
MSE	Mean Squared Error
NMF	Non-negative Matrix Factorization
NSED	Normalized Squared Euclidean Distance
(OR)PCA	(Online Robust) Principal Component Analysis
PDF	Probability Density Function
PGURE	Poisson–Gaussian Unbiased Risk Estimator
SNR	Signal–to–Noise Ratio
SPM	Scanning Probe Microscopy
(S)TEM	(Scanning) Transmission Electron Microscopy
STM	Scanning Tunneling Microscopy
SVD	Singular Value Decomposition
SVT	Singular Value Thresholding

### Mathematical Symbols

$x$	Scalar variable
$\mathbf{x}$	Vector variable
$\ \mathbf{x}\ _1$	$l_1$ -norm of $\mathbf{x}$
$\ \mathbf{x}\ _2$	$l_2$ -norm of $\mathbf{x}$
$\langle \mathbf{a}   \mathbf{b} \rangle$	Inner product of $\mathbf{a}$ and $\mathbf{b}$ ( $= \sum_i a_i b_i$ )
$\mathbf{X}$	Matrix variable
$\mathbf{X}^T$	Transpose of $\mathbf{X}$
$\text{Tr} [\mathbf{X}]$	Trace of $\mathbf{X}$ ( $= \sum_i X_{ii}$ )
$\ \mathbf{X}\ _F$	Frobenius norm of $\mathbf{X}$



---

$\ \mathbf{X}\ _*$	Nuclear norm of $\mathbf{X}$
$\mathbf{A} \circ \mathbf{B}$	Inner product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbb{E}\{X\}$	Expectation of the random variable $X$
$\text{Var}\{X\}$	Variance of the random variable $X$
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean $\mu$ and variance $\sigma^2$
$\mathcal{P}(\lambda)$	Poisson distribution with parameter $\lambda$
$P(A B)$	Probability of the event $A$ given event $B$
$\mathcal{L}(\mathbf{X} \theta)$	Likelihood of the observations $\mathbf{X}$ given the parameter(s) $\theta$
$\hat{\theta}$	Estimator of the parameter(s) $\theta$
$\arg \min_x f(x)$	The argument $x$ which gives the minimum value of $f(x)$
$\mathbf{r}(t)$	Position of a particle at time $t$
$\langle \mathbf{r} \rangle$	Ensemble average of $\mathbf{r}(t)$
$\bar{\mathbf{r}}$	Temporal average of $\mathbf{r}(t)$



# Chapter 1

## Introduction

Ever since its invention, the development of the transmission electron microscope (TEM) has been driven by the desire to observe and characterize materials at the level of single atoms. Writing in 1939, von Ardenne predicted that “sooner or later the ultramicroscopy technique will be able to reveal single atoms and their distribution in the object plane” [1]. Although von Ardenne was the first to propose the scanning transmission electron microscope (STEM), it was the pioneering work led by Albert Crewe in developing the modern STEM that provided the first direct images of single atoms [2–4]. Recent advances have seen the introduction of aberration correctors, which have combined sub-ångstrom image resolution with single-atom spectroscopy in the latest instruments [5]. Another notable hardware development has been increasingly sensitive electron detectors, making it possible to image radiation-sensitive biological specimens with unprecedented resolution [6]. Together these improvements in instrumentation have made (S)TEM a powerful tool for materials characterization at the atomic scale. Structural, chemical, optical and electronic properties can now be mapped and linked directly to first-principles calculations using density functional theory (DFT), thereby bridging the gap between experiment and atomistic modelling [7, 8].

To fully realize the capabilities of (S)TEM characterization, these hardware developments must be matched by innovations in software, particularly in processing and analyzing the large, multi-dimensional datasets that can now be acquired. Significant progress has been made in recent years in applying machine learning techniques to electron microscopy data, opening up new ways of interpreting and understanding the wealth of information the instrument provides. It is vital that the applications of computational imaging are explored further because it provides the robust, objective and reproducible methods necessary for working with large and noisy data [9]. Developing an automated pipeline from data acquisition all the way through to visualization will not only uncover new materials insights, but also drive the design and optimization of experiments, ensuring data is collected in an informed and efficient manner.

One of the most exciting prospects for the modern electron microscope concerns the time-resolved imaging of dynamics at the atomic scale. In its various guises, electron microscopy can provide insights into the behaviour of materials over a substantial range of timescales, from the femtosecond imaging of chemical bond breaking to the millisecond observations of diffusion and surface reconstructions [10]. For STEM imaging in particular, it is the potential applications at longer timescales that are most relevant, since ultrafast imaging typically requires conventional TEM illumination conditions with nanosecond electron pulses [11, 12]. Of course, time-resolved imaging at the atomic scale is not without its difficulties. Issues with image noise, electron beam damage, and identifying and tracking atoms typically limit investigations to stable crystalline systems or to short sequences of only a few atoms. These are not typically sufficient to examine atomic dynamics in detail, thus highlighting the potential role of computational imaging in fully exploiting the temporal applications of the electron microscope.

The capabilities of STEM for studying atomic motion were first recognized in the earliest findings from Crewe's laboratory [13–15]. Atomic diffusion at surfaces plays a pivotal role in many areas of science, not least in epitaxial growth, Ostwald ripening and heterogeneous catalysis [16–18]. Most models of diffusion are based on idealized, crystalline surfaces, but there is an increasing body of evidence suggesting that diffusion in heterogeneous environments departs from the classical picture of Brownian motion, giving rise to so-called *anomalous diffusion* [19]. In much the same way that fluorescence microscopy has revolutionized the study of molecular transport inside cells [20, 21], STEM has the potential to uncover a wealth of previously unknown information about diffusion at the atomic scale.

Notwithstanding the desire to understand fundamentals such as surface diffusion, the continuing efforts in nanotechnology towards the fabrication of ever smaller devices is a major factor in driving the development of aberration-corrected electron microscopy. From transistors to sensors to catalysts, at the nanoscale, the interactions and behaviours of individual atoms are integral to the optical, electronic and catalytic properties of the system. To ultimately control these properties thus requires a full understanding of the local structure with the highest possible resolution. A prime example is the field of heterogeneous catalysis, which combines high-surface area supports with metal nanoparticles for highly efficient and selective catalysts [22]. Common catalysts include the expensive Pt-group metals, which are typically deposited as nanoparticles over a substrate to maximize the available surface area for the reaction.

However, many important catalytic reactions take place at single sites on the surface of nanoparticles, indicating that there is still a significant amount of redundant and expensive metal that plays no part in the reaction [23]. Recent research has shown that spatially-isolated single atoms [24–26] and nanoclusters [27, 28] offer the potential for

highly selective and efficient catalytic reactions, while reducing the amount of the metal catalyst required to the bare minimum. Optimizing this selectivity requires a detailed understanding of the interaction between catalyst and industrially-relevant substrates, which include functionalized materials such as graphene oxide (GO) [29] or graphitic carbon nitride [30]. These materials are low-cost and straightforward to fabricate, and controlling the functionalization to tune the dispersion of the metal atoms enables the optimization of the selectivity and efficiency of the catalyst system [31, 32]. The importance of understanding the distribution and character of these functional groups means aberration-corrected STEM has a significant role to play in understanding the structure of single atom heterogeneous catalysts. Furthermore, the recent development of the environmental STEM raises the prospect of using electron microscopy to track the behaviour of single atoms during catalytic reactions [33–36], thus providing new insights into catalysis.

The aim of this thesis is twofold: first, to develop a computational framework for time-resolved STEM imaging at the atomic scale; and second, to apply these methods to explore and understand the dynamic behaviour of single atoms on technologically-important surfaces. The next chapter provides a general overview of aberration-corrected STEM and its applications, with a focus on single atom imaging. In Chapter 3, the role of machine learning in electron microscopy is discussed, and two novel analysis techniques for time-resolved imaging are presented. Chapter 4 builds on the theme of temporal analysis to explore the challenges in identifying atoms and tracking their motion over time.

Having developed an analytical toolbox for time-resolved STEM, Chapters 5–7 highlight example applications of the various techniques. Chapter 5 is an investigation of the anomalous diffusion of copper adatoms on graphene oxide, whilst Chapter 6 presents results on the structure and behaviour of small copper clusters on GO. Copper clusters have been identified as a low-cost alternative to Pt-group catalysts [37], and these two chapters assess the structure and stability of single atoms and clusters on an industrially-relevant substrate. A separate study of atom dynamics on a crystalline surface is described in Chapter 7. This explores the dynamic behaviour of the silicon surface under the electron beam alongside the beam-induced motion of Pt adatoms, and demonstrates the applicability of the computational imaging methods to both crystalline and disordered substrates. Finally, concluding remarks and possible directions for future research are given in Chapter 8.



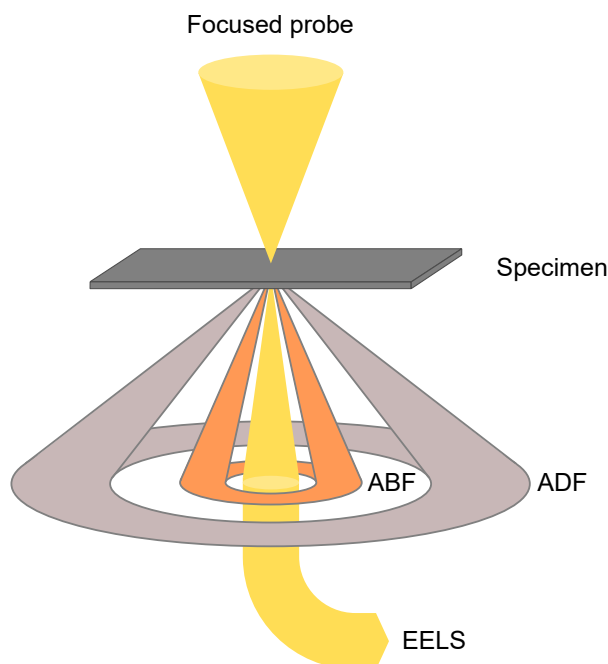
## Chapter 2

# Single atom electron microscopy

### 2.1 Introduction

Electron microscopy is, in its various forms, a widely-used tool for studying the structure and properties of materials at scales beyond the reach of optical microscopy. Resolutions of 0.5 Å are now possible with the latest generation of aberration-corrected instruments [5], while the interaction between the imaging electrons and the specimen also enables chemical characterization with excellent sensitivity [38]. The wealth of information provided by the electron microscope makes it an ideal tool for studying materials with atomic resolution, and also forms the basis of the idea of “multi-dimensional electron microscopy” [39, 40], which seeks to combine different signals to uncover behaviours and properties that any single mode would not detect.

A cornerstone of multi-dimensional electron microscopy is the scanning transmission electron microscope (STEM), with its sensitivity to both light and heavy elements, imaging intensity that scales with atomic number, and simultaneous acquisition of multiple spectroscopic signals for studying bonding, composition and resonance properties. This chapter introduces the STEM and its various operating modes, as well as the developments that have enabled single atom sensitivity. Section 2.3 briefly discusses some of the alternative techniques for direct atomic-scale imaging for comparison with STEM. In the last part of the chapter, the major challenges faced when performing single atom microscopy with STEM are outlined, with a particular focus on time-resolved imaging for observing the dynamic behaviour of atoms at this scale.



**Figure 2.1:** Schematic of the STEM setup, with a focused probe and two detectors: annular bright-field (ABF) and annular dark-field (ADF). Here the central beam passes through to an EELS detector, but this can be replaced with a detector for bright-field (BF) imaging.

## 2.2 Scanning transmission electron microscopy

Transmission electron microscopy (TEM) uses a beam of accelerated electrons to illuminate a thin specimen. These electrons interact with the material as they pass through, where the exiting beam is focused and magnified onto a detector to form an image [41]. The smaller wavelength of accelerated electrons compared to photons makes the resolution of a TEM significantly higher than an optical microscope, and the range of possible interactions between electrons and the sample enables many types of signals to be recorded for both structural and compositional investigation.

Scanning transmission electron microscopy differs from conventional TEM in that a focused electron probe is scanned pixel-by-pixel across a region of the specimen, rather than illuminating the entire region at once (Figure 2.1) [42]. The two techniques are related by the principle of reciprocity [43], and in STEM the resolution and image contrast is controlled by the probe-forming lens and aperture before the specimen, and the geometry of the detector placed after it. By using different detectors it is possible to separately record and investigate unscattered, elastically and inelastically scattered electrons [44]. Annular detectors allow the undiffracted beam to pass through to a



detector for electron energy loss spectroscopy (EELS), enabling simultaneous analysis and characterization [45].

One of the important advances of Crewe and coworkers in the development of atomic-resolution STEM was the use of a high-brightness electron source, the field emission gun (FEG) [2]. The FEG offers a low energy spread of emitted electrons, which is important for both imaging and spectroscopy (see below), while the high brightness maximizes the available probe current for a given probe size, thus maximizing the signal that can be acquired.

The magnetic lenses used in the electron microscope suffer from imperfections, or aberrations, that degrade their performance and the ultimate resolution of the microscope. Particularly important are spherical and chromatic aberrations. In the former case, electrons travelling at high angles to the optical axis are focused too strongly by the lens, while in the latter case electrons with different energies are focused by differing amounts. An energy-filtering slit (or monochromator) can be used to reduce the effects of chromatic aberration at the expense of the probe current, but for the case of spherical aberration it was the development of computer-controlled aberration correctors in the 1990s that marked a turning point in the capabilities of STEM [46–48]. Aberration correctors involve breaking the rotational symmetry that causes spherical aberration, and enable the formation of atom-sized probes with an achievable resolution of ca. 0.5 Å [49]. Alongside this improvement in resolution, focused probes facilitate atom-by-atom spectroscopy, thus revealing not only the position of single atoms but also their identity and bonding with unprecedented resolution [50–52].

### 2.2.1 Bright-field and annular bright-field imaging

In bright-field (BF) and annular bright-field (ABF) imaging, the detectors lie within the cone of illumination, and much of the transmitted intensity comes from electrons that have experienced little or no scattering. Coherent BF imaging can be used to determine crystallographic information about the specimen through diffraction and phase contrast from overlapping diffraction discs [42]. In an ABF imaging setup, the central portion of the beam does not fall onto the detector (Fig. 2.1) [53], which reduces the sensitivity to the phase of the waves scattered by the object. ABF has been shown to be particularly sensitive to light elements such as oxygen and lithium, and has also been used to obtain the first images of hydrogen atomic columns in metal hydride systems [54, 55].

### 2.2.2 Annular dark-field imaging

Annular dark-field (ADF) imaging further suppresses the phase influence of the scattered waves by averaging features over the detector plane [42], and collects electrons scattered beyond the typical Bragg scattering angles of BF imaging. Nellist & Pennycook showed that for a large annular detector the observed intensity depends only on the probe intensity and the fraction of electrons scattered out to the appropriate angles [56]. The ADF image is thus described as a convolution between the STEM probe intensity and an object function  $O$ . The ADF intensity at the probe position  $\mathbf{r}_0$  is then:

$$I(\mathbf{r}_0) = O(\mathbf{r}_0) * |\psi(\mathbf{r}_0)|^2 \quad (2.1)$$

where  $\psi$  is the electron wavefunction. The object function  $O$  determines the intensity fraction scattered by each atom. The incoherency makes ADF imaging directly interpretable, since there is no longer any possibility of contrast reversal as in BF-STEM and conventional TEM imaging. This interpretability is a major factor in the popularity of the technique for atomic-scale materials characterization.

In the incoherent ADF imaging mode the signal intensity scales approximately with atomic number as  $Z^{1.5-2}$ , as electrons passing close to the nucleus are deflected to high angles. This is akin to Rutherford scattering, although effects such as screening of the nucleus reduce the  $Z$  exponent. Importantly, the intensity scaling with atomic number means that ADF images can be directly interpreted and analyzed to collect quantitative data for atom identification and counting [57–59]. The exact value of the exponent will also depend on the geometry of the ADF detector, and here it is noted that the terms high-angle ADF (HAADF) and medium-angle (MAADF) are commonly used in the literature. Generally, HAADF refers to an inner detector angle of ca. 80 mrad or more, while MAADF images may use an inner angle of 30 mrad. The smaller inner angle enhances the contrast of lighter elements such as carbon, which are weak electron scatterers. In this work the general term, ADF, is used throughout; actual detector settings are given where relevant.

While ADF-STEM images are convenient to interpret directly, careful consideration must be taken to consider all of the interactions between the sample and the electron probe. Phonons and thermal motion of atoms in a crystal will displace them slightly from their ideal positions, leading to diffuse scattering of the electrons. It is this diffuse scattering that breaks the image coherence in the  $z$ -direction parallel to the optic axis, since the instantaneous atomic displacements are random [60]. An additional problem with ADF for characterization is that when a crystal is orientated along a zone axis, the electrons will preferentially “channel” down the atomic column, binding tightly to the atoms and possibly leading to higher recorded intensity [61].

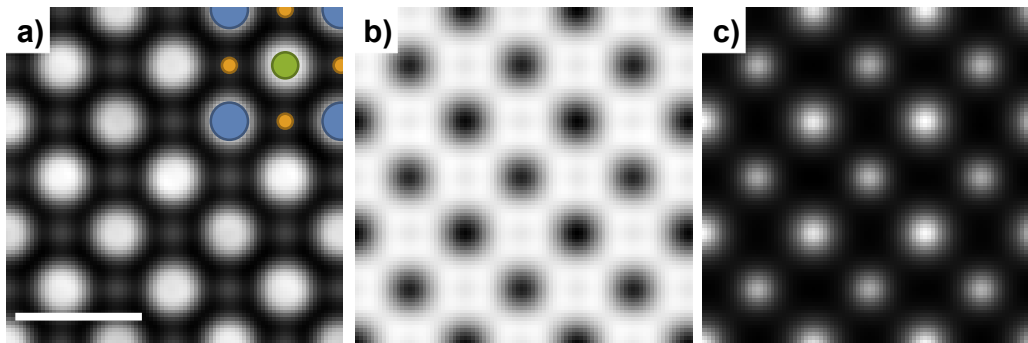
### 2.2.3 STEM image simulation

The development of the theory of electron scattering and electron-sample interactions make it possible to simulate the image produced by an electron microscope, incorporating effects such as aberrations, thermal diffuse scattering and more. By simulating images one is able to compare experimental results with models constructed using other information such as DFT calculations, or even to solve the structure itself by iterating the model until the simulation and experiment match.

There are two well-used methods for (S)TEM image simulation: the Bloch wave method [61], and the multislice method [62]. Of the two, the latter is arguably more flexible in its approach of slicing the structure into thin slices and propagating the electron wave from slice to slice. The Bloch wave method on the other hand is typically a much more demanding calculation, although advances in computing resources mitigate this to some extent, and it is particularly useful for separating out the contributions of different structural features to the final image. Unless otherwise stated, the approach taken in this thesis is the multislice method, using the “Dr. Probe” software package [63].

STEM image simulation software is able to incorporate the effects of aberrations on the electron probe for comparison with experimental results. As well as aberrations, the shape of the electron probe also depends on the size of the electron source (which is not a point source, otherwise the probe current would be zero), and a large source size will limit the image resolution. The typical approach is to convolve the simulated image with a function representing the effective source size [64, 65].

Figure 2.2 is a demonstration of the multislice simulation approach, showing the image contrast produced by BF, ABF and ADF imaging modes applied to the same sample (the perovskite  $\text{SrTiO}_3$ ). This simulation uses an aberration-corrected probe with an accelerating voltage of 300 kV and incorporates thermal effects using the “frozen-phonon” approximation. The differences between the two bright-field modes is clear, as the removal of the central portion of the beam in ABF leads to both contrast reversal and an improved contrast of the light oxygen atomic columns (Fig. 2.2b). The 80 mrad ADF detector in Figure 2.2c is referred to by some in the literature as high-angle ADF (HAADF); this is to distinguish it from smaller medium-angle ADF (MAADF) detectors. The main effect of changing the inner angle is to change the exponent  $\alpha$  in the  $Z^\alpha$  intensity relationship, with larger inner angles such as the 80 mrad HAADF detector (Fig. 2.2c) having a higher  $\alpha$ . The general term ADF is used throughout the rest of this work, and the specifics of each simulation or experiment are given to clarify the results obtained.



**Figure 2.2:** Simulated aberration-corrected STEM images of a 25 nm thick sample of  $\text{SrTiO}_3$ , viewed down the (001) direction. The probe convergence semi-angle was 25 mrad, and the accelerating voltage was 300 kV. **(a)** BF detector (0–5 mrad). **(b)** ABF detector (12–25 mrad). **(c)** ADF detector (80–250 mrad). Positions of Sr (blue), Ti–O (green) and O (orange) atomic columns are marked. Scale bar: 0.5 nm.

#### 2.2.4 Spectroscopy in the STEM

The majority of electrons passing through a thin sample are elastically scattered and so do not lose energy before arriving at the detector. Some electrons do lose energy, however, through inelastic scattering, and as a result contain information about the sample response to the accelerated electrons over a large range of energy losses. These responses can include quasi-elastic phonon scattering, plasmon scattering, or the ionization of inner-shell electrons [44]. The amount of energy lost will depend on the elements and oxidation states within the material, and so the inelastic electrons can be used for spectroscopic analysis in conjunction with the standard modes of image formation. One method is to filter out electrons of a specific kinetic energy before the detector, meaning the image formed depends on the energy loss and therefore, indirectly, the composition of the sample. This is known as energy-filtered TEM (EFTEM) [41].

Instead of filtering out specific electron energies, an alternative is to record a spectrum of energy losses, which will have features again dependent on the sample composition. This is known as electron energy loss spectroscopy (EELS), and as shown in Figure 2.1, the use of annular detectors in STEM means that EELS and ADF imaging can be carried out simultaneously. The low-loss region of the energy spectrum can, for example, be used to determine optical properties of the system, whilst at higher energy losses the ionization edges can be used for identifying elements and understanding the bonding, coordination and band structure of the specimen [52]. Crucially, the focused probe facilitated by aberration correctors provides the ability to map elemental and chemical properties of a specimen at very high spatial resolution, and be matched closely with the ADF image. The low energy spread of the FEG as an electron source is also an important factor in EELS imaging.

As well as the absorption spectroscopy of EELS, further mapping can be conducted using emission spectroscopy, namely recording the energy of emitted X-ray photons resulting from the electron-sample interaction. This is known as energy-dispersive X-ray spectroscopy (EDS or EDX) [66, 67]. Both EDS and EELS have lower scattering cross-sections than ADF-STEM, so dwell times must be longer (or probe currents higher) for quantification and characterization using these spectroscopic methods. This can cause problems with noise when imaging beam-sensitive materials, or with temporal resolution when attempting to analyze dynamic behaviour in which the sample is changing over time.

### 2.2.5 Electron tomography

Transmission electron microscopy is by its very nature a two-dimensional imaging technique, with each image being a projection of a 3D object. For a successful, complete characterization of a material, extracting and understanding the 3D structure is often crucial. Reconstructing a 3D structure from 2D projections or slices is known as *tomography*, and its applications are by no means limited to electron microscopy; tomography is also widely used in medical contexts with X-ray imaging and magnetic resonance imaging (MRI). Here the discussion is limited to (S)TEM tomography, where a series of 2D projections are recorded by rotating the sample about one or more tilt axes, and a processing algorithm applied to reconstruct the 3D object.

A number of tomographic reconstruction methods have been developed for both general and EM-specific applications, and two particular approaches stand out in STEM imaging. The compressed sensing (CS) approach of Leary et al. utilizes powerful aspects of mathematical theory that enable the robust recovery of undersampled signals using a small number of carefully-selected observations [68]. The CS technique not only reduces many of the reconstruction artifacts of other algorithms, but the need for a smaller number of projections vastly improves the time necessary for acquiring a tilt series (important for time-resolved tomography) and also reduces the total electron dose received by the sample, which is crucial for beam-sensitive specimens such as zeolites, polymers and biological materials [69].

The second technique of interest is the discrete tomographic approach of van Aert et al. [70], which utilizes the quantitative nature of ADF-STEM imaging to count the number of atoms in a projected atomic column and combine this information with prior knowledge about the crystallographic structure of the material. The number of projections needed for a successful reconstruction is again reduced, bringing benefits to preventing beam damage, whilst resolution down to the level of single atoms is obtained. Quantitative ADF imaging has also been used alongside atomistic modelling

for the reconstruction of nanoparticles in 3D using only a single projection and a set of sensible assumptions about the structure [71, 72].

In contrast to acquiring many projections of a single object around a tilt axis, a common methodology in the biological sciences is to take a single image of many thousands of identical objects frozen onto a support in random orientations. Aligning and clustering the images by similarity provides the necessary information for the 3D structure to be recovered [73]. This is known as single particle cryo-electron microscopy [74], and the main motivation is that biological specimens are much more sensitive to the electron beam and so cannot withstand more than a single exposure, let alone sufficient images for a tilt series. Hardware and software developments have led to significant advances in cryo-EM capabilities [75], and near-atomic-resolution reconstructions are now possible [76].

As well as the tomographic reconstruction methods mentioned above, the stable, focused probes facilitated by aberration correctors enable the use of the reduced depth-of-field to recover information in the  $z$ -direction parallel to the optic axis. By systematically varying the focus of the microscope the specimen is “depth-sectioned” in a fashion similar to optical confocal microscopy. While the resolution in the image plane is often sub-ångstrom, the depth resolution is typically more limited. The technique has, for example, been used to successfully image screw dislocations in gallium nitride [77].

### 2.2.6 In-situ electron microscopy

Electron microscopes typically require near ultra-high vacuum (UHV) conditions to protect the electron source and to make the passage of the high-energy electrons through the column to the detector as smooth as possible. UHV conditions are atypical of the environments most catalysts and nanoscale devices operate in, making it crucial to understand the difference in behaviours observed in the microscope and in real operating environments. An even better solution is to develop methods for conducting reactions inside the electron microscope, otherwise known as in-situ electron microscopy [78].

One approach to in-situ electron microscopy involves the use of a cell sandwiched between two semi-transparent windows and integrated into the sample holder. Liquids or gases can then be introduced into the cell and reactions observed through the transparent windows [79]. The thickness and transparency of the windows and the cell itself is crucial in determining the achievable signal-to-noise ratio and resolution. An alternative approach devised by Gai & Boyes is the environmental STEM (ESTEM), which uses differential pumping to allow gases into parts of the electron column while protecting the electron source [34], thus removing the need for the windows of a liquid

cell. Combined with further developments with heating and cryo-holders to control the temperature and microelectromechanical systems (MEMS) devices to incorporate strain and measure mechanical responses, it is now possible to follow the development of reactions in controlled, relevant environments all the way down to the scale of single atoms [36].

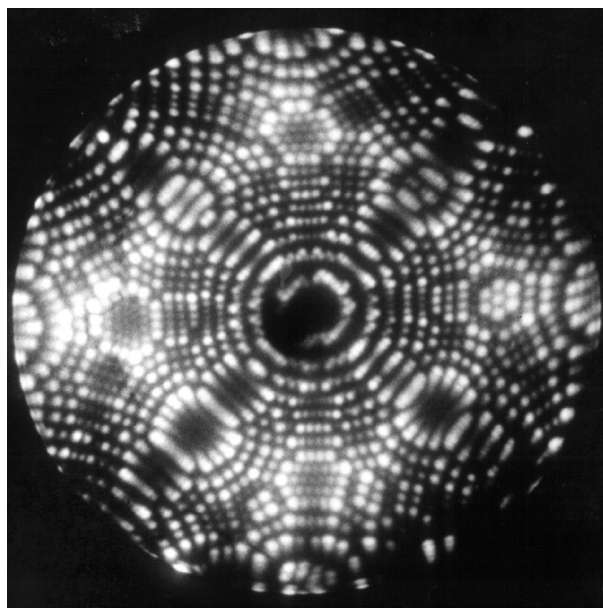
## 2.3 Alternative methods for single atom imaging

Electron microscopy is a powerful technique for imaging and characterization at the atomic scale. However, it is only the development of aberration correctors and related hardware in the last 20 years that has enabled “routine” imaging of single atoms. Prior to these advances, scientists interested in structure of materials at this level relied on a number of other techniques, each with their own advantages and disadvantages. Two of the main techniques for *direct* imaging of single atoms are discussed below, and later chapters will examine how atomic resolution STEM can provide complementary information to these techniques.

### 2.3.1 Field ion microscopy

Field ion microscopy (FIM) was the first technique to “see” individual atoms, a result of work by Müller and Bahadur in 1955 [80]. A sharp needle specimen inside a chamber is positively charged such that the electric field at the tip is very large. The chamber is filled with an imaging gas, typically He or Ne, and at the tip the gas atoms are ionized and accelerated away onto a fluorescent screen, mapping the electric field distribution at the tip. This can then be related directly to the topography of the surface of the tip, and, by maintaining the sample at a low temperature (ca. 50 K), the spatial resolution is sufficient to image surface atoms directly [81]. Figure 2.3 shows an FIM image of the surface of a platinum needle produced using a He–Ne imaging gas, with the sample cooled to ca. 4 K [82]. Increasing the temperature encourages thermally-activated surface diffusion, and FIM has been used to study diffusion in a vast number of systems [83]. If the electric field is increased further, the surface atoms themselves are ionized and evaporate, enabling the specimen to be investigated layer-by-layer. This leads to the related technique of atom probe tomography (APT) for the 3D reconstruction of samples with single-atom resolution [81].

Although FIM and related atom probe techniques are powerful methods for single atom imaging, the high electric fields place a limitation on the samples which can be examined, and surface diffusion studies are typically limited to self-diffusion of refractory metals such as tungsten and platinum. The electric fields can also affect



**Figure 2.3:** FIM image of platinum, with each bright spot corresponding to a Pt atom on the surface of the sample. Reproduced from Ref. 82.

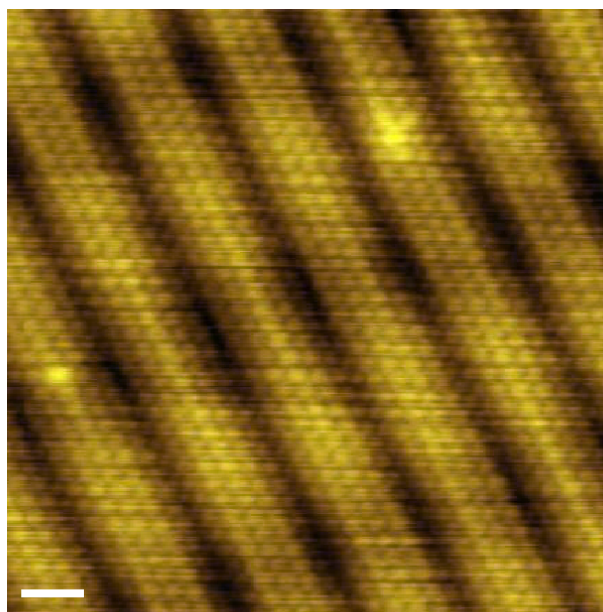
the diffusion itself, while the low temperatures, clean specimens and high vacuum conditions required make FIM unsuitable for in-situ studies.

### 2.3.2 Scanning probe microscopy

Scanning probe microscopy (SPM) uses a physical probe to image the surface of a specimen, and began with the development of the scanning tunneling microscopy (STM) in 1982 by Binnig and coworkers, which famously resolved the Si(111)- $7\times 7$  surface reconstruction with atomic resolution [84]. In STM a conducting tip is placed very near to the surface of interest, and a voltage bias applied. This enables electrons to tunnel between the tip and surface, and the resulting current contains information about the local density of states of the surface at the tip position. By scanning across the surface in a raster fashion, an image of the surface is built up [85]. The related technique of atomic force microscopy (AFM) instead uses the mechanical force between the tip and the surface, so does not require a conducting specimen [86]. Recently it was shown that terminating the AFM tip with a carbon monoxide molecule enables the imaging of single carbon atoms inside benzene rings [87, 88].

The spatial resolution of 0.1 nm is comparable to STEM, while a depth resolution of up to 0.01 nm makes SPM ideal for the study of surface and features such as steps and channels [89], provided the specimen is very flat so as not to obstruct the passage of the probe or damage the tip. In Figure 2.4, the ridges present in the reconstructed surface of Au(100) are identified by the changes in intensity [90]. Also present are





**Figure 2.4:** STM image of the reconstructed Au(100) surface. Scale bar: 1 nm. Reproduced from Ref. 91.

characteristic streaking artifacts in the horizontal scan direction, which can arise from sample movement during acquisition.

While typical SPM image acquisition time is a few minutes, with careful vibration control and electronics design, video-rate STM has been achieved and used to directly image surface diffusion [92–94]. The STM has also been used to manipulate individual atoms on surfaces, opening up a number of applications for nanofabrication of devices at the smallest possible scale [85]. Many of the early experiments with SPM required ultra-high vacuum conditions and cooling to very low temperatures. It is now possible to perform experiments in the presence of liquids and gases and at moderate temperatures for in-situ experiments with catalysts [95, 96], although this typically comes at the expense of atomic resolution [97]. The obvious drawback of SPM techniques is that they are inherently surface-only tools, and no information is recorded about the bulk of the sample, in contrast with the projected 3D information that can be obtained with (S)TEM.

## 2.4 What are the challenges for single atom STEM?

The analytical capabilities of the STEM make it ideally suited for exploring and characterizing the properties of small clusters and single atoms [98]. Such investigations are not without their challenges, however, and this section briefly outlines two of the major issues faced in single atom STEM imaging. The first concerns reducing the dose to reduce the effects of the electron beam on the sample, and the corresponding increase

in noise this brings. The second focuses on time-resolved imaging. Atom motion is to be expected in any in-situ study taking place at an elevated temperature, and that is before any electron beam effects are considered. It is therefore necessary to push the temporal capabilities of the STEM to capture atomic motion where possible.

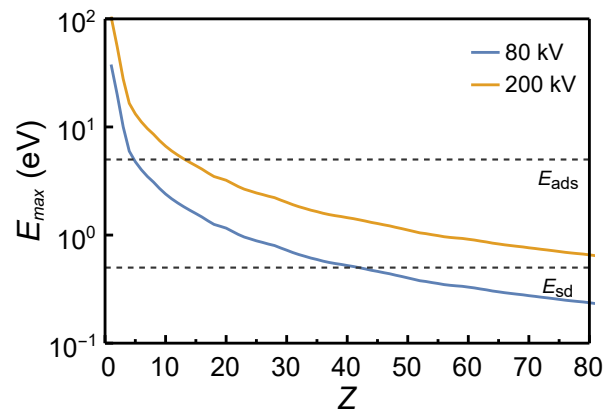
### 2.4.1 Low-dose imaging

Imaging single atoms with ADF-STEM brings with it significant challenges, the most important being the effect of the electron probe itself on the sample under investigation. Individual electrons accelerated through a few tens or hundreds of kV have sufficient energy to knock even a strongly-adsorbed adatom completely off a substrate, and the continual bombardment from the electron probe also leads to localized sample heating [44]. At lower beam energies, the major mechanism of beam damage is changes in the chemical bonding of the specimen as a result of ionization by the incoming electrons. This is known as radiolysis [60].

Reducing the effects of the electron beam is therefore crucial for accurate and robust imaging of beam-sensitive samples. This can be solved in two related ways. First, the electron accelerating voltage can be reduced to minimize the chance of knock-on damage by the probe electrons, which is the main mechanism of damage above a few keV. The maximum energy transferred in a collision between an accelerated electron and an atom is given by the expression [99]:

$$E_{\max}(U) = \frac{2eU(eU + 2m_e c^2)}{m_n c^2} \quad (2.2)$$

where  $U$  is the accelerating voltage,  $m_e$  is the mass of an electron and  $m_n$  is the mass of the atomic nucleus. If the binding energy  $E_d$  of an atom is less than  $E_{\max}$  then the electrons have sufficient energy to displace the atom out of a crystalline lattice. Crucially, adsorption energies of atoms on surfaces are typically lower than bulk cohesive energies, while surface diffusion barriers can be as small as 0.05 eV [100]. The energy of the electron beam must therefore be reduced accordingly if the effects of these collision processes are to be avoided or mitigated [101]. Figure 2.5 plots this as a function of atomic number for two different beam energies, and highlights that even 80 kV electrons have sufficient energy to encourage the surface diffusion of many first-row transition metals. For imaging 2D materials such as graphene, beam energies of 60 kV or less are necessary [51]. Although the reduction in energy reduces the electron wavelength and therefore theoretical resolution of the microscope, the limiting factor in resolution is typically the finite source size and aberrations, so this change is negligible. The development of aberration correctors in particular has made the idea of low-voltage “gentle STEM” a reality for routine single atom imaging [102, 103].



**Figure 2.5:** The maximum energy transferred parallel to a surface between an accelerated electron and an atomic nucleus for two typical STEM voltages. Typical values for adsorption energies ( $E_{\text{ads}}$ ) and surface diffusion barriers ( $E_{\text{sd}}$ ) are marked with dashed lines, and are taken from Ref. 101.

The expression in Equation 2.2 assumes a head-on collision between the electron and the nucleus, but in reality most electrons are only weakly scattered by the nucleus. The average energy transferred therefore depends on the scattering cross-section of the atom, and so the second mechanism for mitigating electron beam effects is to limit the dose received by the sample, since this reduces the chance of an electron–atom collision of sufficient energy and leading to knock-on damage or radiolysis. Organic and polymeric specimens in particular are susceptible to severe radiation damage at doses corresponding to  $1 e^-$  scattered per atom or less, making atomic resolution imaging in such samples very difficult indeed [44, 104]. Minimizing the electron dose can be achieved by reducing the probe current (typically  $10^{-11}$  A in STEM), the probe dwell time per pixel (typically  $10^{-6}$  s), or both. More complicated methods for total dose reduction per image are discussed in Chapter 3, whereby the image is subsampled such that only a fraction of the pixels are in fact exposed to the electron beam, reducing the electrons interacting with a given area of the specimen. The missing observations are then filled in during post-processing using CS methods.

Even though scattering cross-sections in ADF-STEM are considerably higher than in EELS and EDS, only a tiny proportion of electrons are scattered through a sufficient angle to be detected by the annular detector [105]. While the electron dose received by the sample at a given STEM probe position may be  $1000 e^-$ , the number of electrons recorded on the ADF detector could be 100 times smaller. At this level, careful treatment of the counting statistics and noise is very important. The arrival of electrons at a detector is governed by Poisson statistics, and although at higher doses this can be well-approximated by a normal distribution (which is often much simpler to model mathematically when processing or analyzing the images), at such low counts this approximation is less valid. Further noise can be introduced during the readout process from the detector, convolving the Poisson noise with Gaussian noise and giving rise to

a mixed noise model. This distinction between noise models is particularly pertinent for accurate noise removal and atomic-scale characterization with Z contrast imaging, and is discussed in more detail in Chapters 3 and 4.

### 2.4.2 Time-resolved imaging

Acquiring a series of STEM images to provide temporal resolution alongside atomic spatial resolution followed not long after Crewe's first STEM experiments, with the installation of a video camera to record the signal from the ADF detector [106]. Experiments in recent years using aberration-corrected STEM have led to new insights in atom motion on surfaces [107], dynamics of dopant atoms in graphene [108, 109], bulk diffusion of dopant atoms [110], and reversible reaction dynamics [111], with the limiting factors including noise, beam-induced changes, and the rate at which the scan coils can move the probe across the specimen.

Temporal resolutions of up to 10 frames per second (fps) have been achieved, but clearly rapid STEM imaging relies on reduced pixel dwell times, resulting in the same problems with low-dose imaging discussed above. There are also hardware limitations with the rate at which the scan coils in the STEM are able to scan in a stable manner [44], akin to the problems faced in video-rate SPM. This can be mitigated by using coarse pixel sizes; while picometre precision of atomic positions can be obtained by using small pixel sizes, if the interest lies in ångström-scale atomic motion, then the precision needed for atom localization can be reduced; each atom need only be a few pixels across. This idea is discussed further in Chapter 4. A second reason for coarse pixel sizes takes into account the shape and size of the electron probe. Even after aberration correction, typical probes have a full-width half-maximum (FWHM) of ca. 60 pm [60]. Selecting a pixel size less than this value means that a large portion of the probe intensity falls outside of the pixel being observed. The probe tails can thus have a significant effect on the sample, possibly causing atomic motion or damage that cannot be recorded. An example of this effect is investigated in Chapter 7.

Even with coarse pixel sizes, sensitive detectors and robust noise-removal algorithms, the problems of sufficient signal and scan coil stability means that STEM is probably limited to a temporal resolution of  $10^{-6}$  s per pixel, equivalent to about 100 fps assuming a  $128 \times 128$  pixel field-of-view. While smarter scanning patterns and improved hardware may in future improve this resolution, it is unlikely that STEM will achieve the nanosecond timescales at which atomic diffusion takes place [112]. STEM therefore captures the “before” and “after” states of an atom moving between stable sites, reinforcing the need to minimize any unnecessary electron beam effects that might obscure the underlying behaviour.

To image events at improved temporal resolution, one must turn to the parallel illumination of TEM for higher frame rates. For example, new direct electron detectors are capable of acquisition rates of 1000 fps or more with excellent sensitivity and noise characteristics [113]. Separately, the field of ultrafast electron microscopy pioneered by Ahmed Zewail uses rapid pulses of electrons to image materials at the femtosecond ( $10^{-15}$ ) timescale [10]. This stroboscopic technique makes it possible to image atomic motion, resonance behaviour and chemical bonding with unprecedented resolution that, crucially, is well within the reach of *ab initio* molecular dynamics calculations [114]. The method can also be extended further with the tomographic techniques discussed above for “four-dimensional electron microscopy” [115, 116].

## 2.5 Summary

This chapter has outlined the characterization capabilities of the scanning transmission electron microscope. A particular focus is the multiple signals that can be acquired with sub-ångstrom resolution, each providing new insights into the properties of materials at the atomic scale. Image simulation methods provide a direct link between quantitative ADF imaging and atomistic modelling techniques, and electron tomography enables the reconstruction and analysis of materials in three dimensions. The combination of spatial and spectroscopic imaging modes makes STEM an attractive characterization tool compared to other microscopy methods such as SPM or FIM, but there are still challenges surrounding the beam–specimen interaction to overcome in order to fully exploit the temporal capabilities of the microscope. The next chapters establish a number of methods to address these challenges, combining computational imaging methods with machine learning and statistical analysis to develop a framework for time-resolved imaging of single atoms using STEM.



## Chapter 3

# Low-rank methods for electron microscopy

### 3.1 Introduction

Electron microscopy now incorporates a wide range of techniques for both imaging and characterizing materials. In addition to spatial information in two-dimensional imaging and three-dimensional tomography, further data dimensions arise from video microscopy (temporal), EELS and EDS (spectral), and diffraction (reciprocal space). As the volume of data collected in microscopy experiments grows, there is an ever-increasing need to process the datasets in a manner that extracts the key information content. Since manual analysis is not practical for large datasets, machine learning methods are a key step to analyzing and interpreting the information. These methods must be able to handle not only the size and complexity of the data, but also robustly deal with noise and missing data, which are common problems in electron microscopy. Objective, reproducible analysis is only possible through rigorous consideration of the image acquisition process, noise characteristics, and other experimental constraints and complications. Of particular interest for in-situ and time-resolved microscopy studies are algorithms capable of real-time analysis, offering the potential for denoising, processing and analysis at the microscope.

This chapter demonstrates the applications to electron microscopy of low-rank machine learning methods, which exploit correlations in large datasets to extract the most important features. Building on a number of commonly-used methods for multi-dimensional analysis, a robust low-rank denoising algorithm is introduced\*, and example applications in time-resolved electron microscopy are used to demonstrate the performance

---

\*The work on singular value thresholding for denoising time-resolved image sequences (Section 3.3) was published in Ref. 117 in collaboration with R.K. Leary and P.A. Midgley.

and broad applicability of the algorithm. The second half of the chapter focuses on online algorithms for denoising and for subsequent analysis such as background subtraction, which are key steps in the workflow for processing large datasets. The potential for real-time processing in electron microscopy using these online algorithms is assessed with several pertinent examples, including the problem of missing or corrupted data.

### 3.2 Matrix factorizations in electron microscopy

In electron microscopy, significant attention has been focused in recent years on a family of matrix factorizations and decompositions, which fall under the umbrella of unsupervised machine learning methods [118]. Matrix factorizations typically seek to approximate the original data with a lower-dimensional representation. Dimensionality in this sense has a specific meaning, and refers to the dimension spanned by the columns (or rows) of the matrix, otherwise known as the matrix *rank*. Seeking a low-rank representation of a dataset enables otherwise hidden correlations to be exploited, thus aiding the interpretation of the experimental observations.

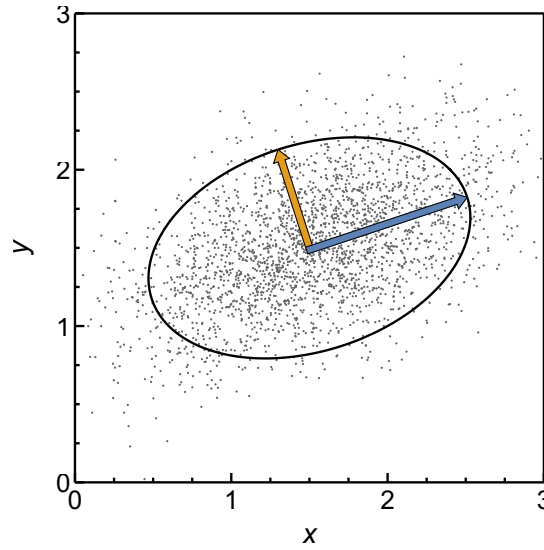
Let  $\mathbf{X}^0$  be an  $m \times n$  data matrix of  $m$  features and  $n$  observations. Matrix factorization methods seek to decompose the data into factor matrices  $\mathbf{A}$  and  $\mathbf{B}$ , according to [119]:

$$\begin{aligned}\mathbf{X}^0 &= \mathbf{A}\mathbf{D}\mathbf{B}^T \\ &= \sum_{i=1}^r \lambda_i \mathbf{a}_i \mathbf{b}_i^T\end{aligned}\tag{3.1}$$

where  $\mathbf{A}$  is a  $m \times r$  matrix,  $\mathbf{B}$  is a  $r \times n$  matrix, and  $\mathbf{D}$  is a  $r \times r$  diagonal scaling matrix. Here  $r$  represents the rank of the matrix, and when  $r \ll m, n$  the matrix is said to be low-rank. In cases where  $r$  is large it is often desirable to instead find a low-rank matrix approximation where  $\hat{\mathbf{X}} \approx \mathbf{X}^0$ ; this will be explored in detail in Section 3.3. The rank 1 variables  $\mathbf{a}_i \mathbf{b}_i^T$  are known as *components* of the matrix  $\mathbf{X}^0$ , with  $\mathbf{a}_i$  typically referred to as *factors* and  $\mathbf{b}_i$  as *loadings*.

Solving Equation 3.1 relies on appropriate constraints, otherwise there are an infinite number of combinations of  $\mathbf{A}$  and  $\mathbf{B}$ . The next section discusses some of the more common methods for applying constraints, and their respective applications in electron microscopy.



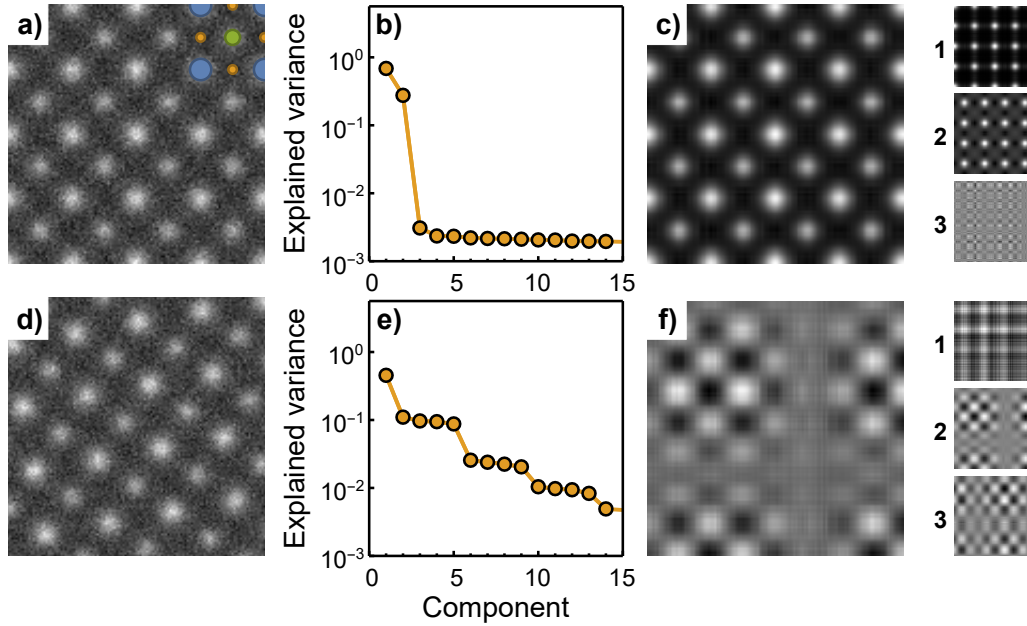


**Figure 3.1:** PCA applied to 2500 random samples drawn from a bivariate Gaussian distribution with mean (1.5, 1.5). The two arrows are the eigenvectors of the covariance matrix of the points, scaled by the eigenvalues.

### 3.2.1 Principal component analysis

Principal component analysis (PCA) imposes an orthogonality constraint on Equation 3.1 to transform the data into a set of linearly independent components [120]. The first principal component is chosen to explain as much of the variance of the observations as possible. Subsequent components are then chosen to explain as much of the remaining variance whilst being orthogonal to the previous components. It is typically conducted by applying an eigenvalue decomposition to the covariance matrix of the data, since each orthogonal principal component is an eigenvector of the covariance matrix. Figure 3.1 is a simple example using random samples drawn from a bivariate Gaussian distribution. The principal components highlighted give the new orthogonal axes for the transformed data. Alternatively, one can apply a singular value decomposition (SVD), defined as  $\mathbf{X}^0 = \mathbf{U}\Sigma\mathbf{V}^T$ .  $\mathbf{U}$  and  $\mathbf{V}$  are both orthogonal matrices ( $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$ ). The values along the diagonal of  $\Sigma$  are denoted as  $\sigma_i$ , and are known as the singular values of the matrix  $\mathbf{X}^0$ . They are sorted in descending order, and are a measure of how much of the data variance is explained by each principal component.

PCA is often used for denoising electron microscopy datasets by filtering out the unwanted noise components, which are typically uncorrelated, and leaving behind a low-rank representation of the dataset [121]. This is demonstrated in Figure 3.2a, where Gaussian noise has been added to a simulated ADF-STEM image of  $\text{SrTiO}_3$ . The scree plot in Figure 3.2b is derived from the singular values of the noisy image, and shows that ca. 99.7% of the variance in the image can explained by a linear



**Figure 3.2:** (a) A simulated ADF-STEM image of SrTiO<sub>3</sub> is corrupted by a small amount of Gaussian noise. The atomic columns are marked: Sr (blue), Ti-O (green) and O (orange). (b) The scree plot obtained by applying PCA to the image shows most of the variance in **a** is explained by the first two principal components. (c) The image reconstructed from the first two components (shown right), which correspond to the Sr and Ti-O columns. The third component is mostly noise. (d–f) Rotating the image by 10° destroys the low-rank property.

combination of just two components, which in this case are the heavy Sr atomic columns and the lighter Ti-O atomic columns. The remaining principal components contribute very little to the image, and essentially explain the uncorrelated Gaussian noise. By removing these components, the image can be reconstructed as a low-rank approximation (Fig. 3.2c). The low-rank property is critical, as Figure 3.2d–f shows that even a small rotation off-axis destroys the correlation between matrix rows (or columns). Similar problems would be encountered for aperiodic features such as vacancies or dopant atoms. To overcome this problem, one instead takes a stack of similar images, for example video frames or slices along the energy axis of an EELS spectrum, and flattens the images into vectors. The assumption is that while variations along rows or columns within in image may be considerable, correlations along the third axis are likely to be stronger and maintain the low-rank property. Such an approach is suitable for most practical applications. Finally, although PCA is an effective denoising method when the low-rank assumption holds, selecting the number of components to retain is not always as clear as Figure 3.2b, and robust methods for automating this selection are necessary; this is the focus of Section 3.3.

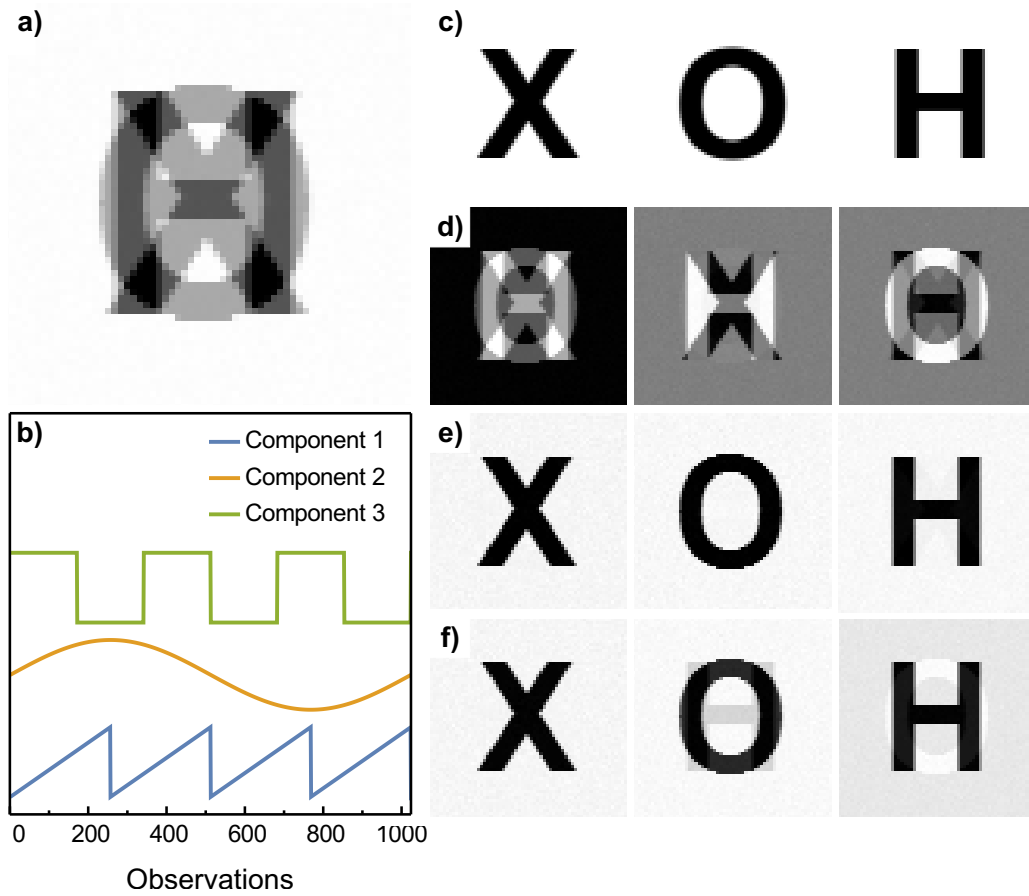
### 3.2.2 Physical interpretability

For denoising applications PCA is a straightforward and effective tool, since the components themselves are not necessarily of interest. However, in many cases in electron microscopy, the components themselves are extremely useful, as they correspond to individual features of the dataset. Figure 3.2 presented a very simple application, with the first two principal components corresponding approximately to the Sr and Ti–O atomic columns.

Realistically, most electron microscopy datasets are not as simple as this example, and often exhibit principal components that are difficult to interpret in a straightforward physical manner, even though they accurately explain the variance of the dataset. One particular problem for electron microscopy is that PCA does not prevent negative values from appearing in the factors or loadings, which is problematic considering the positive values inherent to count data. Non-negative matrix factorization (NMF) seeks to overcome this issue by factorizing the data into two matrices with no negative elements, instead of the orthogonality constraint imposed by PCA. NMF has been successfully applied to the analysis of surface plasmons with EELS [122, 123], as well as for decomposing scanning electron diffraction data for local orientation mapping of crystals [124, 125].

The non-negativity constraint typically leads to improved interpretability of the NMF components over PCA, as shown in Figure 3.3. Here, three  $64 \times 64$  pixel images (Fig. 3.3c) have been flattened (or vectorized) to form a  $4096 \times 3$  matrix. This is multiplied by a mixing matrix made up of three different signals (Fig. 3.3b) to give a stack of vectorized images with dimensions  $4096 \times 1024$ . The mean image of these observations is shown in Figure 3.3a, after a small amount of Poisson noise has been added. The observation matrix is first decomposed using PCA, which struggles to accurately recover the three component images (Fig. 3.3d); although these components do explain the variance of the dataset, they are not recognisable compared to the original images. NMF performs much better on the other hand, as can be seen in Figure 3.3e, by constraining the components to be strictly non-negative.

Instead of non-negativity, independent component analysis (ICA) enforces the statistical independence of each component, which has proven particularly useful for spectroscopic analysis in electron microscopy [126–128]. An example of ICA is shown in Figure 3.3f, where two of the recovered components are similar to the NMF results. The failure to fully isolate the second and third components is a result of the small vertical correlation between the letters O and H.



**Figure 3.3:** (a) The mean image from a stack of 1024 images formed by the mixing matrix of the three signals in **b** and the three component images in **c**. A small amount of Poisson noise has also been added to the stack. (d) The first three components recovered by PCA. (e) The three components recovered by NMF. (f) The three components recovered by ICA.

### 3.2.3 Challenges

Despite the improvements in interpretability gained by NMF and ICA, the problem of selecting the appropriate number of components to describe the dataset remains. Too few components and the reconstructed result is underfitted; too many components and the model is overfitted to describe the noise. Usually a clear “elbow” in the PCA scree plot is sufficient to identify the point at which the remaining components are mostly describing the noise, but in many situations it is less obvious. Section 3.3 investigates automated, objective methods for selecting this cut-off.

Furthermore, although PCA is a very common method for noise removal prior to subsequent analysis, the method is itself susceptible to the specific noise conditions due to a least-squares assumption in the solver. For Gaussian noise, where the noise variance is independent of the underlying pixel intensity, this least-squares assumption is valid. However, for conditions such as Poisson noise this no longer holds, since the variance

of each data point now depends on the value of that data point. Weak but important features are often hidden in the factorization results due to the noise bias. A simple yet effective Poisson normalization method was presented by Keenan & Kotula [129, 130], which can improve the extraction of physically-significant features using PCA. For scenarios where the electron or photon counts are extremely low (typically  $<10$ ), Salmon et al. have developed an effective Poisson PCA algorithm [131]. As well as Poisson noise, PCA can also be affected by large, sparse errors caused by missing pixels or X-ray spikes, and this problem is addressed in depth in Section 3.4.

Finally, the size of datasets can restrict the tools available for processing and analysis. Even on modern computer hardware, matrix factorizations and decompositions for large datasets can be extremely costly both in terms of floating-point operations and memory requirements. In particular, the SVD algorithm used for PCA and many other methods exhibits poor scaling behaviour for large datasets. As microscopy hardware develops and datasets grow in size and complexity, overcoming these computational challenges is critical for interpreting experimental results. Incremental SVD algorithms have been developed for solving the memory restriction by only considering a fraction of the dataset at a time [132], and a related technique is discussed in Section 3.4.

### 3.2.4 Tensor methods

The use of matrix methods in electron microscopy assumes that the multi-dimensional dataset can be flattened into a 2D matrix for analysis. In practice this is a straightforward process; for example, an EELS dataset will have a spectral axis and two spatial axes, and the latter two can be combined to give a single navigation axis. However, as new dimensions are added (such as time, or complementary signals such as EDS), or as the size of the datasets increases, careful consideration of these operations is necessary to ensure the analysis remains both physically meaningful and computationally feasible.

Tensor decompositions are a natural extension to the matrix methods presented in this work. Operating on the dataset in its original multi-dimensional form requires new algorithms, since the definition of a tensor rank is less clear, but can bring additional benefits for both data storage and the computational complexity of the analysis [119]. Constraints such as non-negativity are straightforward to incorporate [133], and tensor methods also offer additional flexibility such as enforcing non-negativity in one dimension and enforcing independence or sparsity in another. Datasets can also be linked together for scenarios where only some components are common between datasets [134], as might arise in combined EELS/EDS experiments where the cross-sections of some elements can be very different.

### 3.3 Singular value thresholding for denoising

Figure 3.2 demonstrated that a noisy dataset can be cleaned by exploiting the low-rank structure within the data. Using only a subset of the principal components gives a low-rank approximation to the noisy dataset, which then gives a closer representation of the original low-rank data prior to corruption by noise. Restating this problem more formally, given a corrupted observation  $\mathbf{Y}$  of a low-rank matrix  $\mathbf{X}^0$ , the goal of low-rank matrix approximation is to recover  $\mathbf{X}^0$  as accurately as possible. A natural approach is to find the optimal solution to:

$$\arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \text{rank}(\mathbf{X}) \quad (3.2)$$

where  $\mathbf{X}$  is the decision variable and  $\lambda$  is a regularization parameter [135].  $\|\mathbf{Y} - \mathbf{X}\|_F^2$  represents the square of the Frobenius norm, i.e. the sum of the squared differences of the matrix elements,  $\sum_{ij} |Y_{ij} - X_{ij}|^2$ . Stating the problem in this way imposes a low-rank constraint on the estimated matrix  $\mathbf{X}$ , whilst ensuring  $\mathbf{X}$  is also a good fit of the observations,  $\mathbf{Y}$ .

As with many other machine learning approaches, the selection of appropriate parameters can be a significant challenge. The previous section showed how the number of principal components used to approximate a dataset with PCA needs careful consideration to avoid either under- or overfitting. Robust, objective methods for parameter selection are therefore critical to developing and applying machine learning to ever-larger and more complex datasets, where manual parameter tuning is both time-consuming and subjective. This section also discusses the parameter selection issue, and it is addressed for the problem of denoising under specific noise conditions, although many of techniques involved have found applications in other areas of mathematics and signal processing [136].

#### 3.3.1 Nuclear norm minimization

In practice, optimization of the rank function in Equation 3.2 is an intractable problem. However, Candès and Recht demonstrated a powerful approach involving minimization of the nuclear norm of the matrix as a convex approximation to the rank function [135]. Nuclear norm minimization seeks to approximate  $\mathbf{X}^0$  by an optimal low-rank solution  $\hat{\mathbf{X}}_\lambda$  according to:

$$\hat{\mathbf{X}}_\lambda = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \|\mathbf{X}\|_* \quad (3.3)$$

where  $\|\mathbf{X}\|_*$  is the nuclear norm of  $\mathbf{X}$ , which is defined as the sum of the singular values of a matrix,  $\sum_i \sigma_i$  [135]. Cai et al. showed that a solution to Equation 3.3 can be found in a computationally attractive manner using the soft singular value

thresholding (SVT) operation [137], according to:

$$\hat{\mathbf{X}}_\lambda = \text{SVT}_\lambda(\mathbf{Y}) = \mathbf{U} \mathcal{S}_\lambda(\Sigma) \mathbf{V}^T \quad (3.4)$$

where  $\mathcal{S}_\lambda$  is the operator, and for each singular value  $\sigma_i$ :

$$\mathcal{S}_\lambda(\sigma_i) = \max[\sigma_i - \lambda, 0] \quad (3.5)$$

The soft thresholding operator contrasts with the typical hard thresholding approach of PCA, which retains only those principal components with singular values above a threshold  $\lambda$  [130]. Equation 3.5 instead reduces all the singular values towards zero by a fixed amount.

For practical application to images and image sequences, a modification to Equation 3.4 is made. Instead of applying a single threshold value to all singular values, a weighted threshold can be applied on the basis that larger singular values correspond to more important image features, and so should be reduced by a smaller amount [138]. Defining the weighted nuclear norm of  $\mathbf{X}$  as  $\|\mathbf{X}\|_{w,*} = \sum_i w_i \sigma_i$ , where  $w_i \geq 0$  is a weight assigned to the singular value  $\sigma_i$ , a solution is now sought for:

$$\hat{\mathbf{X}}_\lambda = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \|\mathbf{X}\|_{w,*} \quad (3.6)$$

where the parameter  $\lambda$  has been incorporated into the weighted nuclear norm. The approach taken in Ref. 138 uses weights  $w_i$  in the order  $0 < w_1 < \dots < w_n$  (based on the fact that the singular values of a matrix are always sorted in descending order  $\sigma_1 > \sigma_2 > \dots > \sigma_n$ ). This ensures the use of soft singular value thresholding remains valid [139]. Applying an exponential weighting scheme instead reduces the computational complexity compared to the original approach of Ref. 138. Letting  $\sigma_{\max} = \max[\Sigma]$ , the exponentially-weighted SVT operator incorporating the parameter  $\lambda$  is:

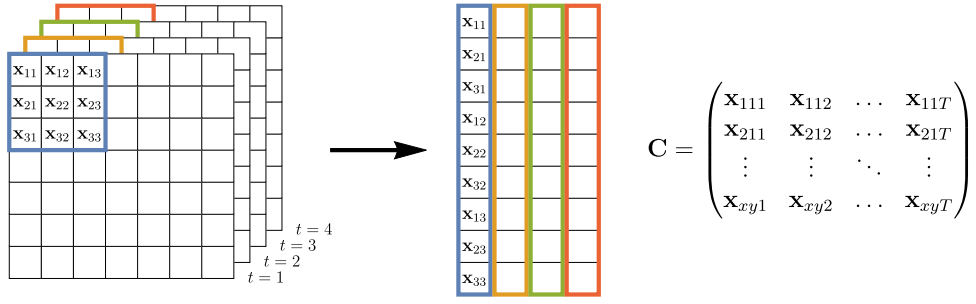
$$\mathcal{W}_\lambda(\sigma_i) = \max \left[ \sigma_i - \sigma_{\max} \exp \left( -\frac{\sigma_i^2}{2\lambda^2} \right), 0 \right] \quad (3.7)$$

and the weighted SVT function is:

$$\text{WSVT}_\lambda(\mathbf{Y}) = \mathbf{U} \mathcal{W}_\lambda(\Sigma) \mathbf{V}^T \quad (3.8)$$

### 3.3.2 Application to time-resolved microscopy

In forming a so-called Casorati matrix, whose columns are the vectorized frames from an image sequence, the correlation between frames in the sequence means that such a matrix will be low-rank [140]. This Casorati matrix is then suitable for the nuclear norm minimization treatment described above.



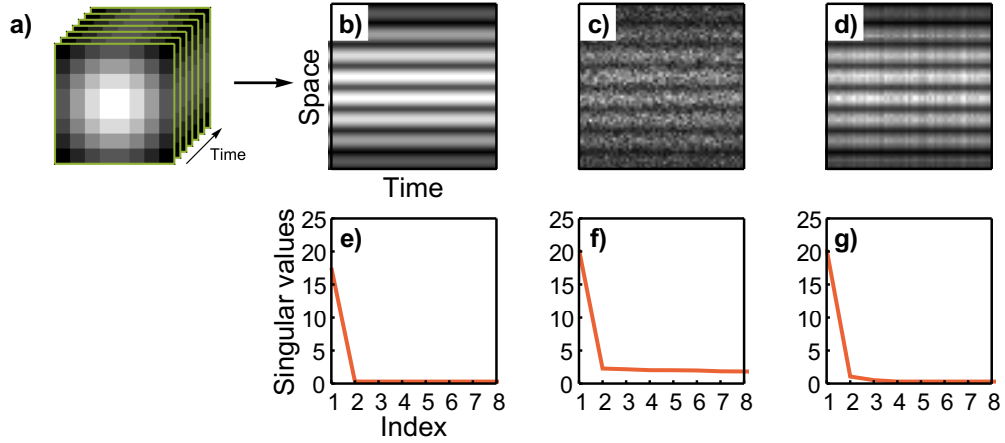
**Figure 3.4:** The construction of a Casorati matrix  $C$  from patches of an image sequence  $x$  with  $t = 1, \dots, T$  frames. Figure reproduced from Ref. 117.

In reality however, the size of the spatial dimension will often exceed the size of the temporal dimension ( $n_x n_y \gg T$ , where  $n_x n_y$  is the number of pixels in each frame and  $T$  is the number of frames), and this may lead to problems with the SVT approach due to limited degrees of freedom [141]. Analyzing the image sequence via a patch-based approach can overcome this problem. The patch-based approach is illustrated in Figure 3.4, whereby a  $3 \times 3$  pixel patch is extracted from each frame and vectorized to form a column of the Casorati matrix  $C$ . This patch-based adaptation means that local rather than global spatio-temporal correlations are considered, giving a more flexible algorithm to cope with variations across image frames, as well as being straightforward to parallelize for modern computer architectures.

Figure 3.5 shows an example of weighted SVT applied to a Casorati matrix formed by vectorized images of a 2D Gaussian peak (representing, for example, an atomic column). The resulting Casorati matrix is in fact rank 1, as can be seen in the singular value plot in Figure 3.5e. These singular value plots can be interpreted in a similar way to the scree plots of PCA, in which most of the variance in the Casorati matrix (and by extension the original image sequence) can be explained by the first component, and the remaining components only describe the noise in the dataset (Fig. 3.5f). In PCA, these remaining components would be discarded to reconstruct the low-rank approximation, whereas in nuclear norm minimization all the components are retained, but their influence on the result is reduced according to the soft thresholding operation (Fig. 3.5d,g).

When applying the SVT algorithm to video denoising, it is important to ensure that the Casorati matrix formed by patches from consecutive frames is indeed low-rank [141]. In cases where motion between frames is significant this assumption may no longer hold. This is illustrated in Figure 3.6b, where the unaligned patches from a simulated image sequence do not form a low-rank Casorati matrix. To overcome this problem, the motion of a patch through the sequence can be estimated, and the trajectory information used to align the individual patches and hence reduce the rank of the





**Figure 3.5:** (a) A stack of frames constituting a sequence in which each image contains the same Gaussian peak. (b) The corresponding rank 1 Casorati matrix formed by the full sequence. (c) The Casorati matrix corrupted by noise. (d) The Casorati matrix after exponentially-weighted SVT ( $\lambda = 1.9$ ). (e–g) Plots of the singular values of the Casorati matrices in (b–d). Figure reproduced from Ref. 117.

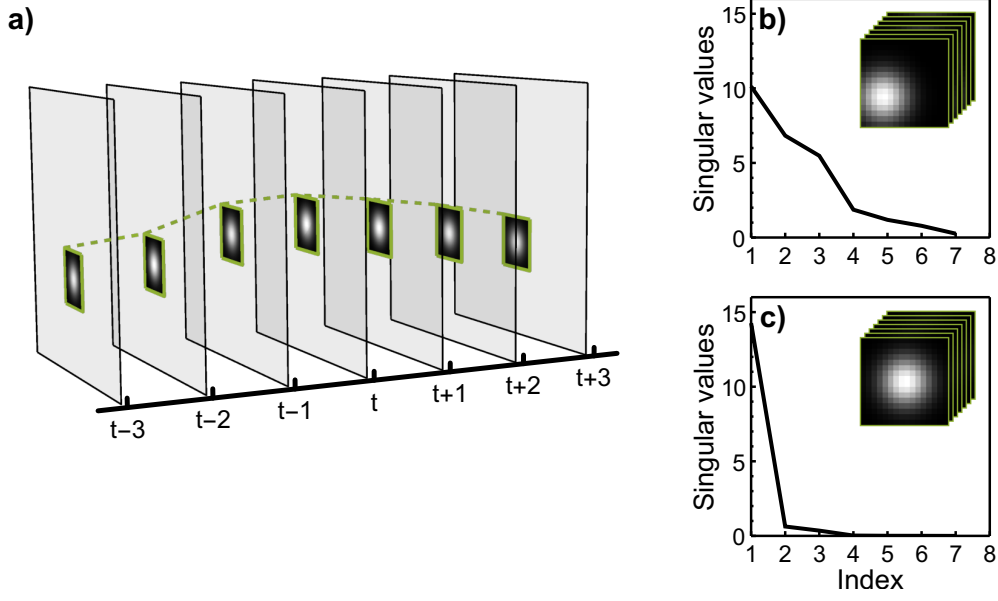
Casorati matrix (Figure 3.6c). Given the position of a patch in frame  $t$ , the motion from frame  $t$  to  $t + 1$  can be estimated by searching the local neighbourhood in frame  $t + 1$  for the most similar patch. An adaptive rood pattern search (ARPS) method is used to reduce the computational cost of the motion estimation [142]. A simple median filter is also used to improve the result of the ARPS step, although all SVT operations are performed on the unfiltered data.

### 3.3.3 Optimal parameter selection

Having developed a weighted SVT function for denoising matrices and image sequences (Eq. 3.8), the next problem is determining the appropriate amount of thresholding to apply, as highlighted in Section 3.2.3. If  $\lambda$  is too large then noise is insufficiently removed from the matrix, and if it is too small then a bias will be introduced as the estimated matrix has a lower rank than the original. Robust and automated methods for obtaining the optimum value of  $\lambda$  are of critical importance to handle this tradeoff between a biased model and a model explaining too much of the data variance.

If  $\mathbf{Y}$  is a noisy observation of the matrix  $\mathbf{X}^0$ , and  $f_\theta$  represents a denoising function dependent on the parameter(s)  $\theta$ , then it is desirable to select  $\theta$  such that the difference between the denoised signal  $f_\theta(\mathbf{Y})$  and the ground-truth  $\mathbf{X}^0$  is minimized. Dropping the subscript, a common method is to minimize the mean-squared error (MSE) or risk:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \|f(\mathbf{Y}) - \mathbf{X}^0\|_F^2 \quad (3.9)$$



**Figure 3.6:** (a) Trajectory of a simulated particle through an image sequence. (b) Singular values of the Casorati matrix formed by unaligned patches extracted from the sequence (c) Singular values of the Casorati matrix formed by patches aligned to the reference frame  $t$ . Figure reproduced from Ref. 117.

where  $N$  is the number of elements of the matrix. In practice, the ground-truth  $\mathbf{X}^0$  is not available to optimize this expression, so instead unbiased estimators of the risk can be used such that  $\mathbb{E}\{\text{MSE}\} = \mathbb{E}\{\text{estimator}\}$ . Selecting the appropriate estimator first requires a consideration of the characteristics of the signal acquisition process, since an understanding of the detector and noise statistics is key to maximizing the performance of the denoising algorithm.

Constructing an unbiased estimator therefore requires an accurate treatment of the noise-generating model of the system. In an additive zero-mean Gaussian noise model, where  $\sigma^2$  is the variance of the noise, the noisy matrix  $\mathbf{Y}$  is defined as:

$$\mathbf{Y} = \mathbf{X}^0 + \mathbf{E} \text{ with } \mathbf{E} \sim \mathcal{N}(0, \sigma^2) \quad (3.10)$$

The unbiased estimator of the MSE in this case is known as Stein's Unbiased Risk Estimator (SURE) [141]. A similar Poisson Unbiased Risk Estimator (PURE) can be derived for a Poisson noise model [143], where the noisy observations are drawn from a Poisson distribution,  $\mathcal{P}$ :

$$\mathbf{Y} \sim \mathcal{P}(\mathbf{X}^0) \quad (3.11)$$

For STEM image acquisition, the mixed Poisson-Gaussian noise model was introduced in Chapter 2 as being a more appropriate and more general model by accounting for the counting statistics of electrons arriving at the detector as well as the readout noise

in the detector electronics. Now the noisy matrix  $\mathbf{Y}$  is defined as:

$$\mathbf{Y} = \alpha \mathbf{Z} + \mathbf{E} \quad \text{with} \quad \begin{cases} \mathbf{Z} \sim \mathcal{P}\left(\frac{\mathbf{x}^0}{\alpha}\right) \\ \mathbf{E} \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \quad (3.12)$$

where  $\alpha$  is the gain of the detector, and  $\mu$  corresponds to a detector offset. As  $\alpha \rightarrow 0$  the model corresponds to Gaussian noise (Eq. 3.10), and conversely, by setting  $\alpha = 1$  and  $\mu, \sigma = 0$  the model corresponds to Poisson noise (Eq. 3.11). In Appendix B the work of Ref. 144 is extended to incorporate a non-zero detector offset  $\mu$ , and a Poisson-Gaussian Unbiased Risk Estimator (PGURE) is defined as:

$$\text{PGURE} = \frac{1}{N} \left( \|\mathbf{f}_\lambda(\mathbf{Y}) - \mathbf{Y}\|_F^2 + 2\mu \mathbf{f}_\lambda(\mathbf{Y}) + 2(\alpha(\mathbf{Y} - \mu) + \sigma^2) \circ \partial \mathbf{f}_\lambda(\mathbf{Y}) - 2\alpha\sigma^2 \partial^2 \mathbf{f}_\lambda(\mathbf{Y}) - (\alpha + \mu) \mathbf{Y} + \mu \right) - \sigma^2 \quad (3.13)$$

where  $\circ$  denotes the element-wise product of two matrices. Combining this estimator with the weighted SVT denoising function (Eq. 3.8) gives a denoising algorithm called PGURE-SVT.

In Ref. 144 the authors outlined an empirical method for calculating Equation 3.13 that does not directly evaluate the terms  $\partial \mathbf{f}_\lambda(\mathbf{Y})$  or  $\partial^2 \mathbf{f}_\lambda(\mathbf{Y})$ . The empirical method allows denoising algorithms to be used as “black-box” processes, requiring no knowledge of their partial derivatives. This empirical approach is taken for PGURE-SVT, although a closed-form solution for Gaussian noise with SURE-SVT was derived in [141]\*.

Note that in deriving the expression for PGURE, there is a restriction on the denoising function  $\mathbf{f}_\lambda$ , in that it must be smooth and weakly differentiable, and the further restriction that  $\mathbb{E} \left\{ \sum_n \left| \frac{\partial \mathbf{f}_\lambda}{\partial Y_n}(\mathbf{Y}) \right| \right\} < +\infty$ . Denoising functions which do not meet this requirement include hard thresholding methods, for example applied to Fourier or wavelet coefficients or to singular values. The latter example is important when relating SVT to PCA. Both approaches seek a low-rank approximation to a dataset, but in PCA only those components with singular values above a threshold are retained for reconstruction. An objective method for selecting the optimum number of components would of course be preferable to visual inspection of the scree plot, but the hard thresholding of the singular values means that the unbiased risk estimators presented here cannot be applied directly to PCA.

---

\*The issue of automated parameter selection within the tensor framework was recently addressed for Gaussian noise with unbiased estimators [145], and also more generally using Bayesian methods [146].

### 3.3.4 Noise estimation

While not relying on the ground-truth  $\mathbf{X}^0$ , the PGURE expression assumes that all three parameters of the Poisson-Gaussian noise model,  $\alpha$ ,  $\mu$ , and  $\sigma^2$  are known. This is often not the case in practical applications, and the parameters must be estimated instead. Wavelet methods are often used to estimate the amount of zero-mean Gaussian noise [136], but estimating the parameters for mixed Poisson-Gaussian noise is more complicated [147–149]. The approach taken here follows work in Ref. 149, and is based on the premise that homogeneous regions of each frame of an image sequence can be used to fit equations relating the expectation and variance of a noisy signal. Each frame in the sequence is partitioned into regions of homogeneous variance using a quadtree segmentation procedure [150]. A robust estimate of the mean and variance of each region is then obtained. From the mixed noise model in Equation 3.12:

$$\mathbb{E}\{\mathbf{Y}_i\} = \alpha \mathbf{X}_i^0 + \mu \quad (3.14)$$

$$\text{Var}\{\mathbf{Y}_i\} = \alpha^2 \mathbf{X}_i^0 + \sigma^2 \quad (3.15)$$

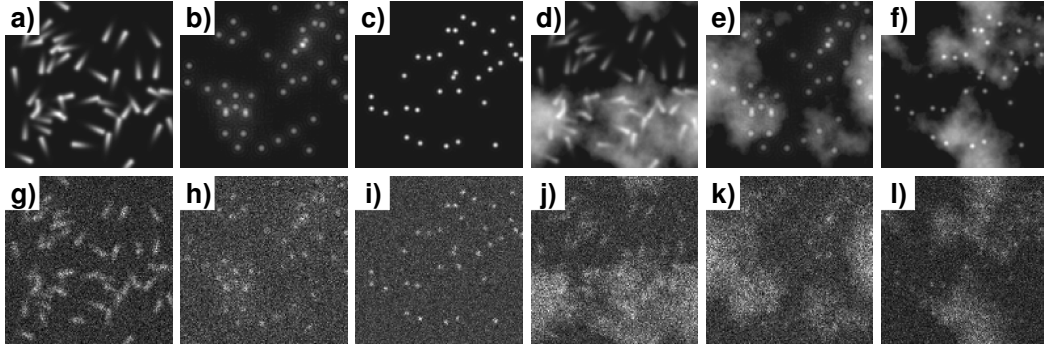
where the subscript  $i$  refers to the  $i^{\text{th}}$  region from the quadtree segmentation, and  $\mathbb{E}\{\mathbf{Y}_i\}$  and  $\text{Var}\{\mathbf{Y}_i\}$  represent the mean and variance of the region  $\mathbf{Y}_i$ . These two identities are combined for Equation 3.16, which shows that a linear regression in  $(\mathbb{E}\{\mathbf{Y}_i\}, \text{Var}\{\mathbf{Y}_i\})$  will provide an estimate of  $\alpha$  and  $(\sigma^2 - \alpha\mu)$ :

$$\text{Var}\{\mathbf{Y}_i\} = \alpha \mathbb{E}\{\mathbf{Y}_i\} + (\sigma^2 - \alpha\mu) \quad (3.16)$$

To complete the noise estimation, the values of the detector offset  $\mu$  and variance  $\sigma^2$  must be determined from the intercept,  $(\sigma^2 - \alpha\mu)$ . Often,  $\mu$  can be estimated directly from the image sequence to be denoised as  $\min(\mathbb{E}\{\mathbf{Y}_i\})$ . This will usually be the case in fluorescence microscopy where only a fraction of the space in each image contains signal from the fluorescing sample. This also true in general for any sequence in which regions of the image do not contain any sample (e.g. regions corresponding to vacuum in STEM). If a sufficient background region free of signal from the sample is not available, for example in ADF-STEM imaging of a continuous sample or a substrate that fills the field of view, then the solution is to obtain a separate reference image  $\mathbf{R}$ , using the same acquisition settings, to give the sample-free background, and therefore  $\mu \approx \min(\mathbb{E}\{\mathbf{R}_i\})$ .

### 3.3.5 Simulation studies

To investigate the performance of the PGURE-SVT algorithm and applicability to time-resolved microscopy, three test sequences were taken from a recent study of particle

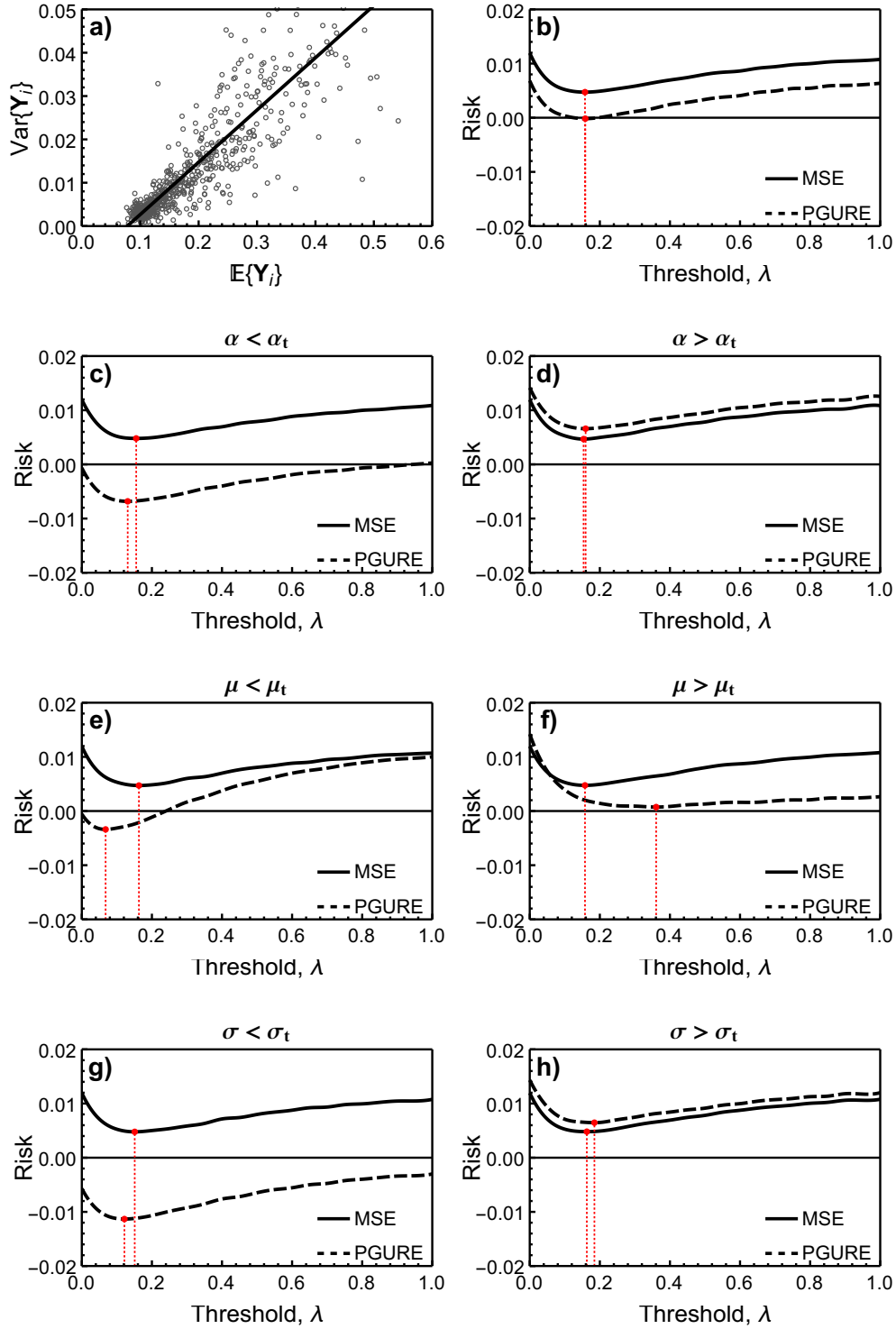


**Figure 3.7:** Example frames from the six test sequences. (a) *Microtubule*, (b) *Vesicle*, (c) *Receptor*, (d) *Microtubule+clouds*, (e) *Vesicle+clouds*, (f) *Receptor+clouds*. (g–l) Noisy versions of the same frames, corrupted by Poisson-Gaussian noise with  $\alpha = 0.1$ ,  $\sigma = 0.1$  and  $\mu = 0.1$ . Figure reproduced from Ref. 117. See also Movies S1–S6.

tracking methods for fluorescence microscopy [151]. Three further sequences were generated by adding a randomly-varying “cloudy” background to the original data. The sequences are deliberately chosen because of the differing types of particle motion, for testing of the ARPS motion estimation procedure. The *Microtubule* sequence exhibits directed motion, *Vesicle* exhibits Brownian (diffusive) motion, and *Receptor* is a combination of directed and Brownian motion. Example frames, normalized to an intensity range between 0 and 1, are shown in Figure 3.7 before and after the addition of mixed Poisson-Gaussian noise. The sequences are available on the accompanying CD (see Movies S1–S6).

We first analyze the relationship between PGURE and the MSE and show how the estimator is a useful tool for optimizing the thresholding parameter, and also show an example of the noise estimation procedure. The *Microtubule* sequence was corrupted with Poisson-Gaussian noise according to the mixed noise model with parameters  $\alpha = 0.1$ ,  $\sigma = 0.1$  and  $\mu = 0.1$ . Figure 3.8a shows the robust estimates of the local mean and variance according to the quadtree segmentation procedure, and the solid line is the fit of Equation 3.16 to these estimates. The slope of the fit in Figure 3.8a therefore corresponds to the estimated value of  $\alpha$ , and the  $y$  intercept to the estimated value of  $(\sigma^2 - \alpha\mu)$ .

Figure 3.8b demonstrates that when the noise parameters are known exactly, both MSE and PGURE predict the same optimal threshold  $\lambda$  (the offset in the  $y$ -axis is irrelevant here). Scenarios where each noise parameter was over- and under-estimated while the others were held constant were then considered, to investigate the effects of shortcomings in the noise estimation procedure. The sensitivity of PGURE to the estimated values of  $\alpha$  and  $\sigma$  is small (Fig. 3.8c,d,g,h), as although there is again a significant offset in the  $y$ -axis, the location of the minima of the MSE and PGURE are relatively similar.



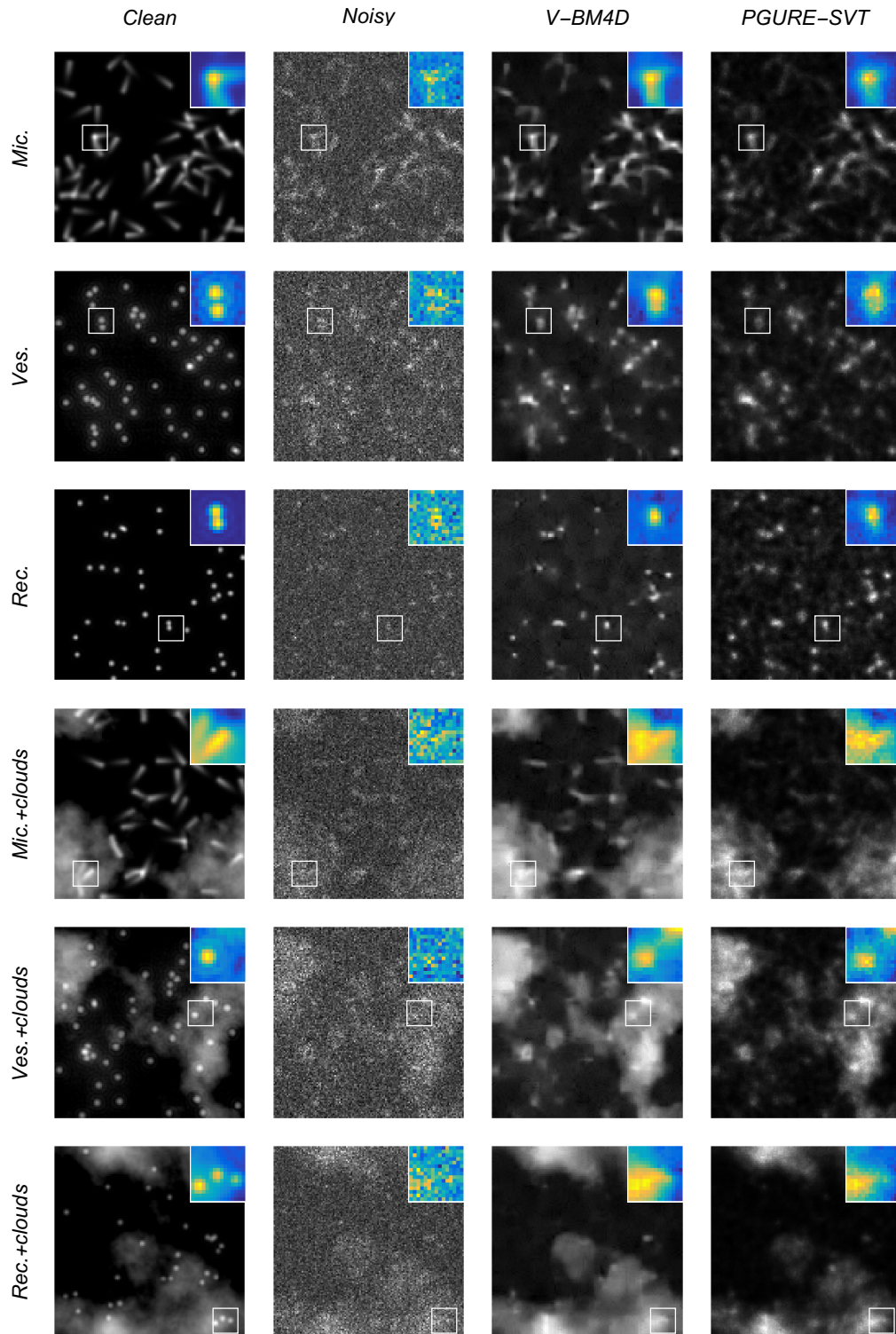
**Figure 3.8:** (a) Robust estimation of the local mean and variance after quadtree segmentation for the *Microtubule* sequence corrupted with Poisson-Gaussian noise with true values  $\alpha_t = 0.1$ ,  $\sigma_t = 0.1$  and  $\mu_t = 0.1$ . (b) Comparison of MSE and PGURE for exact noise parameters. (c, d) For over- and under-estimated  $\alpha$ . (e, f) For over- and under-estimated  $\mu$ . (g, h) For over- and under-estimated  $\sigma$ . The red dotted line indicates the position of the minima of the curves. Figure reproduced from Ref. 117.

Figures 3.8e,f show that PGURE is much more sensitive to the estimate of the detector offset,  $\mu$ . This is particularly important in the case of ADF-STEM images, where a non-zero background due to the presence of a substrate can lead to the poor estimation of  $\mu$ . This will then lead to a discrepancy in the value of  $\lambda$  predicted by MSE and PGURE, resulting in too much or too little shrinkage being applied to the cleaned sequence. This highlights the importance of either an accurate estimate of  $\mu$ , or suitable specimen-free background region(s) within the images to ensure the best performance.

The performance of the PGURE-SVT algorithm is also compared with another patch-based method, V-BM4D, which is widely regarded as the current state-of-the-art in video denoising [152], and incorporates a motion estimation step similar to that used in PGURE-SVT. In this comparison, the default algorithm parameters defined in the source code available from the authors are used [153]. V-BM4D is designed for videos corrupted with zero-mean Gaussian noise, so here it is combined with the Generalized Anscombe Transform for variance stabilization to deal with the Poisson-Gaussian noise model [154]. The algorithm consists of two filtering steps, the first of which involves a hard threshold. As stated in Section 3.3.3, this means that unbiased risk estimators cannot be applied to automate the parameter selection unlike PGURE-SVT. The size of each patch is also limited to  $2^N$  pixels in V-BM4D, whereas PGURE-SVT can be freely adjusted to suit the size of the features in the image sequence.

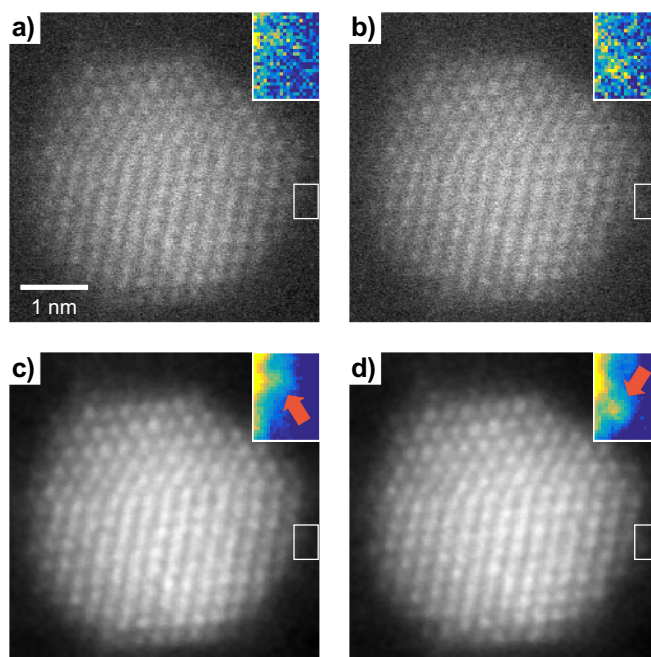
Figure 3.9 shows clean, noisy and denoised frames from each of the test sequences. The insets highlight particular regions of interest in each sequence; for example, two neighbouring particles in the *Vesicle* and *Receptor* sequences. Under the challenging noise conditions, both V-BM4D and PGURE-SVT are able to recover much of the detail from the image sequences. The collaborative filtering and weighted aggregation method applied by V-BM4D creates “blocking” artifacts, with round particles such as in the *Vesicle* sequence having a square shape. The exponentially-weighted shrinkage of PGURE-SVT, on the other hand, results in perceptually smoother denoising that matches the shapes of the particles more successfully.

The perceptual benefits of PGURE-SVT are further highlighted in the three sequences with a cloudy background, where V-BM4D again produces “blocking” artifacts and also washes out much of the intensity variations of the clouds. This is apparent in the inset of the *Vesicle+clouds* sequence, where the washed-out clouds in the V-BM4D image make identification of the particle against the background harder than the comparable PGURE-SVT image. It is noted that many image features appear brighter in the V-BM4D results compared to PGURE-SVT, which is a result of the global shrinkage applied in the weighted SVT step.



**Figure 3.9:** Clean, noisy and denoised frames from six simulated test sequences. The noisy sequences were corrupted by Poisson-Gaussian noise with  $\alpha = 0.1$ ,  $\sigma = 0.1$  and  $\mu = 0.1$ . Figure reproduced from Ref. 117. See also Movies S1–S6.



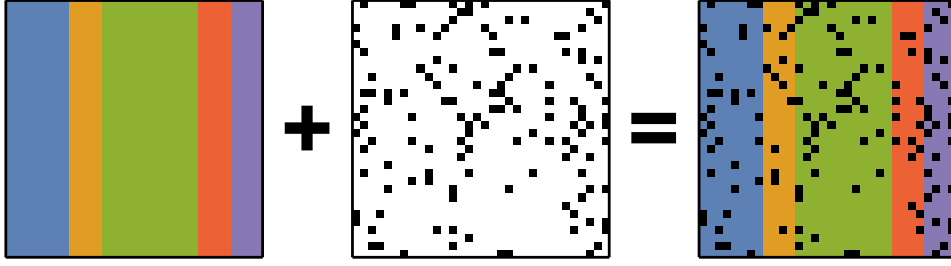


**Figure 3.10:** (a,b) Frames 4 and 5 from an ADF-STEM sequence of a nanoparticle. (c,d) The same frames denoised with PGURE-SVT. The inset images highlight the movement of a single atom along the edge of the nanoparticle (marked by the arrows). Figure reproduced from Ref. 117. See also Movie S7.

### 3.3.6 PGURE-SVT applied to nanoparticle dynamics

The patch-based adaptation of the PGURE-SVT algorithm means that variations in motion, intensity and noise across a frame in an image sequence can be accounted for. Figure 3.9 showed how PGURE-SVT can recover the motion of relatively isolated point sources as found in fluorescence microscopy and single atom STEM imaging. Further applications of the algorithm for recovering the motion of single atoms are presented in Chapters 5–7, but here it is demonstrated how PGURE-SVT can also be applied to more densely arranged image features, such as the periodic structures of crystalline materials imaged with STEM. Of particular interest is the structural rearrangement of single atoms at the surface of a nanoparticle [107], since this motion is typically lost or blurred when noisy image stacks are aligned and averaged with the usual approaches in STEM [155, 156].

Figure 3.10 shows two frames from an image sequence of a nanoparticle on a graphene oxide support (see Movie S7). Aberration-corrected ADF-STEM imaging was performed on an FEI Titan<sup>3</sup> (S)TEM equipped with a CESCOR probe aberration corrector. The image sequence was acquired at 80 kV, with a probe semi-convergence angle of 20 mrad and an ADF detector inner angle of 36 mrad. The beam current was 60 pA and the per-pixel dwell time was 3  $\mu$ s. The measured lattice spacings of the nanoparticle



**Figure 3.11:** Schematic diagram of the robust PCA problem, which combines a low-rank matrix with sparse errors. Figure adapted from Ref. 158.

are consistent with either Au or Ag, but the actual identity is not important here for the purposes of demonstrating the performance of the denoising algorithm. Applying PGURE-SVT to the sequence acquired at 4 fps shows that the denoised results readily reveal atomic-scale surface dynamics; for example, revealing the motion of an atom at the edge of a large nanoparticle as highlighted in the insets (Fig. 3.10c,d).

### 3.4 Robust principal component analysis

The use of a Frobenius norm for the data fidelity term in Equation 3.6 strictly assumes a Gaussian model on the corrupting noise in order for the estimate  $\hat{\mathbf{X}}_\lambda$  to be optimal. As discussed in Section 3.2.3, Poisson noise can cause problems with the interpretation of individual components, although the PGURE-SVT algorithm provides a partial solution by using small overlapping image patches, within which the image intensity, and by extension noise variance, is fairly homogeneous. A more pressing concern is scenarios that lead to more severe corruptions, such as missing data or spikes in intensity caused by X-rays or dead pixels. The Frobenius norm is highly sensitive to such outliers, and in these cases a more robust data fidelity term based on the Kullback-Leibler divergence may be able to deal with arbitrarily large corruptions as well as Poisson noise [157].

An alternative is to relax the low-rank factorization to include a low-rank component  $\mathbf{X}^0$  and a sparse component  $\mathbf{S}^0$ , thus providing a method to account for the severe corruptions, which are assumed to be random and sparsely distributed through the dataset. The noisy observation matrix  $\mathbf{Y}$  is then given by:

$$\mathbf{Y} = \mathbf{X}^0 + \mathbf{S}^0 \quad (3.17)$$

Figure 3.11 shows a schematic diagram showing how the low-rank and sparse components are combined. The method for solving Equation 3.17 was christened *robust* principal component analysis (RPCA) [159], due to the similarities between the low-rank

matrix approximation here and “standard” PCA. The solution involves the following optimization problem for  $\mathbf{X}$  and  $\mathbf{S}$ :

$$\arg \min_{\mathbf{X}, \mathbf{S}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X} - \mathbf{S}\|_F^2 + \lambda \|\mathbf{X}\|_* + \xi \|\mathbf{S}\|_1 \quad (3.18)$$

where  $\|\mathbf{S}\|_1 = \sum_{ij} |S_{ij}|$  is the  $l_1$ -norm of  $\mathbf{S}$ ,  $\lambda$  is the nuclear norm regularization parameter (cf. Eq. 3.3) and  $\xi$  controls the sparsity of the error term,  $\mathbf{S}$ .

Optimization with the  $l_1$ -norm is a well-studied problem in the field of compressed sensing for the recovery of undersampled signals [160]. Provided the signal is sparse in some domain, compressed sensing enables efficient recovery with fewer samples than dictated by the Shannon-Nyquist sampling theorem. In a similar fashion to the nuclear norm approximating the rank function, the  $l_1$ -norm is a useful surrogate for the  $l_0$ -norm, i.e. the number of non-zero components in the data. The presence of two regularization parameters,  $\lambda$  and  $\xi$ , makes automated estimation akin to PGURE-SVT more complicated since the interaction between the parameters is less clear. Further investigation of this interaction is required and it may be the case that a simple risk optimization approach may not be possible for the two-parameter case.

### 3.4.1 Online robust principal component analysis

Numerous algorithms have been developed to solve Equation 3.18 [161, 162]. However, these implementations typically require all the samples and so are unsuited to real-time analysis or to large datasets. Of particular interest in electron microscopy is the online robust PCA (ORPCA) algorithm developed by Feng et al. [163], which updates the low-rank and sparse components as each new observation arrives. This opens up the opportunity for real-time denoising and processing on the microscope, adapting to any experimental changes, rather than processing data *post factum*.

ORPCA is made possible by expressing  $\mathbf{X}$  as the product  $\mathbf{L}\mathbf{R}^T$ , which means an upper bound can be placed on the rank of  $\mathbf{X}$ <sup>0</sup> [164]. Feng et al. incorporated this bound into Equation 3.18\* to give:

$$\arg \min_{\mathbf{L}, \mathbf{R}, \mathbf{S}} \frac{1}{2} \|\mathbf{Y} - \mathbf{L}\mathbf{R}^T - \mathbf{S}\|_F^2 + \frac{\lambda}{2} \left( \|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2 \right) + \xi \|\mathbf{S}\|_1 \quad (3.19)$$

Relating this back to the matrix factorization definition in Equation 3.1, we can see that  $\mathbf{L}$  represents the underlying subspace, or factors, describing the data, and  $\mathbf{R}$  represents the loadings. If  $\mathbf{Y}$  is a dataset with  $m$  features and  $n$  samples, with one sample  $\mathbf{y}_t$  arriving per time instance  $t$ , then Equation 3.19 can be solved in an online manner by

---

\* A similar approach for subspace tracking is used in Ref. 165. Alternative methods for performing online robust PCA have been proposed, e.g. Ref. 166, but so far only the convergence of ORPCA has been investigated theoretically.

first solving for the columns  $\mathbf{r}_t$  and  $\mathbf{s}_t$ :

$$\arg \min_{\mathbf{r}, \mathbf{s}} \frac{1}{2} \|\mathbf{y}_t - \mathbf{L}_{t-1} \mathbf{r} - \mathbf{s}\|_2^2 + \frac{\lambda}{2} \|\mathbf{r}\|_2^2 + \xi \|\mathbf{s}\|_1 \quad (3.20)$$

The underlying subspace  $\mathbf{L}$  is then updated by solving:

$$\arg \min_{\mathbf{L}} \frac{1}{t} \sum_{i=1}^t \left( \frac{1}{2} \|\mathbf{y}_i - \mathbf{L} \mathbf{r}_i - \mathbf{s}_i\|_2^2 + \lambda \|\mathbf{r}_i\|_2^2 + \xi \|\mathbf{s}_i\|_1 \right) + \frac{\lambda}{2t} \|\mathbf{L}\|_F^2 \quad (3.21)$$

This brings significant computational benefits over batch methods that incorporate a singular value decomposition, since the memory required for each iteration is independent of the number of samples,  $n^*$ . The algorithm can be continually updated with new information without the need to re-run over all previous samples, which enables the tracking of any changes in  $\mathbf{L}$  over time. The initial estimate of  $\mathbf{L}$  at  $t = 0$  can be random, as in the original algorithm, or trained on the first few samples of  $\mathbf{X}$  using a QR decomposition with minimal computational impact to the real-time algorithm<sup>†</sup>.

The focus of this work is on the applications of the ORPCA algorithm to denoising time-resolved electron microscopy, but it is worth noting that it is just one example of an online, robust machine learning method. Making a minor adaptation to Equation 3.18 (replacing  $\mathbf{X}$  by  $\mathbf{V}\mathbf{H}$  where  $\mathbf{V}$  and  $\mathbf{H}$  are non-negative) leads to an online NMF method with similar robustness to outliers [167]. This could be particularly beneficial for online EELS and EDS analysis where physical interpretability in the components is key.

### 3.4.2 Stochastic gradient descent

Key to the performance of the ORPCA algorithm is the rate at which the estimate  $\mathbf{L}$  approaches the true subspace, as this dictates the number of samples needed for accurate processing. Equation 3.21 is solved in the original work using either a closed-form or a block coordinate descent (BCD) method. Both of these methods make use of a running total of all the previous samples, which can be computationally detrimental if the dimension  $m$  or rank  $r$  are large. Problems with convergence will also arise if the subspace  $\mathbf{L}$  is changing over time, since the use of an average over all previous samples will mask any changes. Finally, the closed-form solver involves the calculation of a matrix inverse, which can suffer from numerical instabilities.

---

<sup>\*</sup>Memory requirements for batch methods are  $\mathcal{O}(mn)$ . The memory requirements of ORPCA instead depend on an initial estimate of the rank,  $r$ , of  $\mathbf{L}$ , as  $\mathcal{O}(mr)$ . The algorithm complexity is also significantly reduced, from  $\mathcal{O}(nm^2)$  to  $\mathcal{O}(mr^2)$ .

<sup>†</sup>Defined as  $\mathbf{X} = \mathbf{Q}\mathbf{R}$  where  $\mathbf{Q}$  is orthogonal and  $\mathbf{R}$  is an upper triangular matrix. If  $\mathbf{X}$  has rank  $r$ , then the first  $r$  columns of  $\mathbf{Q}$  form the basis of  $\mathbf{X}$ .

To overcome the problems of the pre-existing solvers, a stochastic gradient descent (SGD) algorithm is derived in Appendix C. This uses information from just a single sample to update  $\mathbf{L}$  (Eq. 3.21), enabling the algorithm to adapt much faster to changes in the subspace. If ORPCA is to be used in real-time as data is recorded on the microscope, this adaptability is crucial for dealing with typical experimental instabilities such as sample drift.

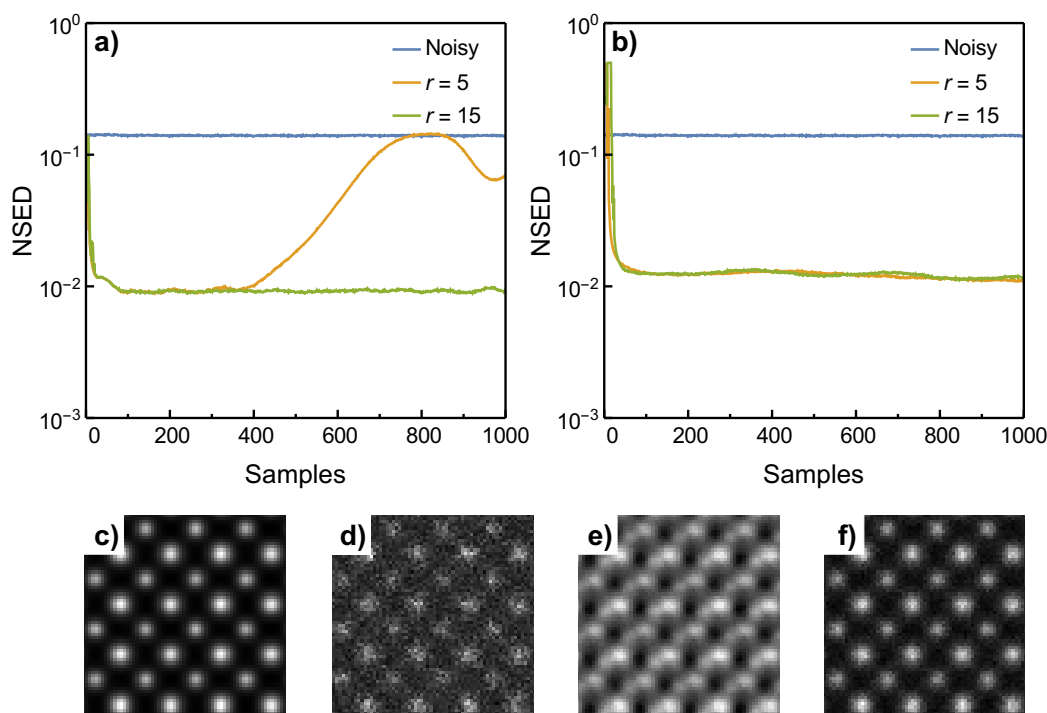
A simulated ADF-STEM image sequence of  $\text{SrTiO}_3$  was used to compare the new SGD method with the BCD method. Poisson-Gaussian noise ( $\alpha = 0.1$ ,  $\mu = 0$  and  $\sigma = 0.1$ ) was added to the image sequence, as well as a gradual sample drift to investigate the ability of both methods to adapt to changes in the low-rank component. There are no sparse errors in this scenario, so setting the sparse parameter  $\xi$  to an arbitrarily large value ensures any effect of the estimated sparse component is negligible. The nuclear norm parameter  $\lambda$  was set to the default value of  $1/\sqrt{m}$  throughout, as per the original paper [163]. The SGD step size or learning rate determines how much information to use from the current sample compared to the previous samples (see Appendix C.3), and a value of 1.1 was found to work well in terms of speed and accuracy with the example datasets here. More advanced methods such as AdaDelta do away with learning rates altogether by relying instead on previous gradients to solve Equation 3.21 [168].

The performance was measured using a normalized squared Euclidean distance (NSED), defined for two images  $\mathbf{X}$  and  $\mathbf{Y}$  as:

$$\text{NSED} = \frac{1}{2} \frac{\|(\mathbf{X} - \bar{\mathbf{X}}) - (\mathbf{Y} - \bar{\mathbf{Y}})\|_F^2}{\|\mathbf{X} - \bar{\mathbf{X}}\|_F^2 + \|\mathbf{Y} - \bar{\mathbf{Y}}\|_F^2} \quad (3.22)$$

This normalized definition is chosen over the conventional MSE or peak signal-to-noise ratio (PSNR) metrics because the shrinkage applied by nuclear norm minimization has the effect of reducing the overall intensity of the image, and the resulting offset can dominate the metric calculation over the actual image features when comparing denoising performance with other methods.

Figure 3.12a shows how the BCD method struggles with the sample drift when  $r$  is small. The recovered image (Fig 3.12e) is blurred as a result of the running average of all previous samples (see also Movie S8). Using a nuclear norm regularizer (Eq. 3.18) means that  $r$  is in fact only an estimate of the true rank of the dataset, and while increasing it improves the flexibility of the algorithm for fitting the model to the data (one must of course be wary of overfitting), it also comes with an increased computational overhead. The SGD method on the other hand displays a much lower sensitivity to the initial rank estimate (Fig. 3.12b,f). Although the BCD method displays a faster convergence at the start of the sequence, the adaptability and improved computational performance of the SGD method is beneficial for large datasets. In



**Figure 3.12:** Online denoising of a simulated  $\text{SrTiO}_3$  ADF-STEM image sequence with Poisson-Gaussian noise and sample drift. The performance of the (a) BCD and (b) SGD methods are shown as a function of the number of samples, for two different initial rank estimates. (c) Frame 800 from the original sequence, (d) after the addition of Poisson-Gaussian noise, (e) recovered by the BCD method ( $r = 5$ ), and (f) recovered by the SGD method ( $r = 5$ ). See also Movie S8.

particular, the fact that SGD is less sensitive to the initial rank estimate (similar behaviour is also seen for the sensitivity to  $\lambda$  and  $\xi$ ) makes SGD an attractive method when very little is known about the dataset prior to acquisition. Smart adaptations to SGD have also been developed that may improve the rate of convergence to match that of BCD; these are discussed in Appendix C.4, although further testing is necessary to understand how to select the appropriate parameters for these approaches.

### 3.4.3 Denoising large datasets

New developments in direct electron detectors, offering high efficiency and low readout noise, have garnered a significant amount of interest in the electron microscopy community [113]. Significant decreases in electron dose and increases in the achievable frame rate are particularly attractive to in-situ and time-resolved experiments [169], as well as cryo-EM research [76, 170].

However, with the increased frame rate comes a significant increase in the size of the acquired data. Taking the K2 Summit camera (Gatan, Inc.) as an example, the full frame size is 16 million pixels, and the maximum frame rate at this size is 400 fps.

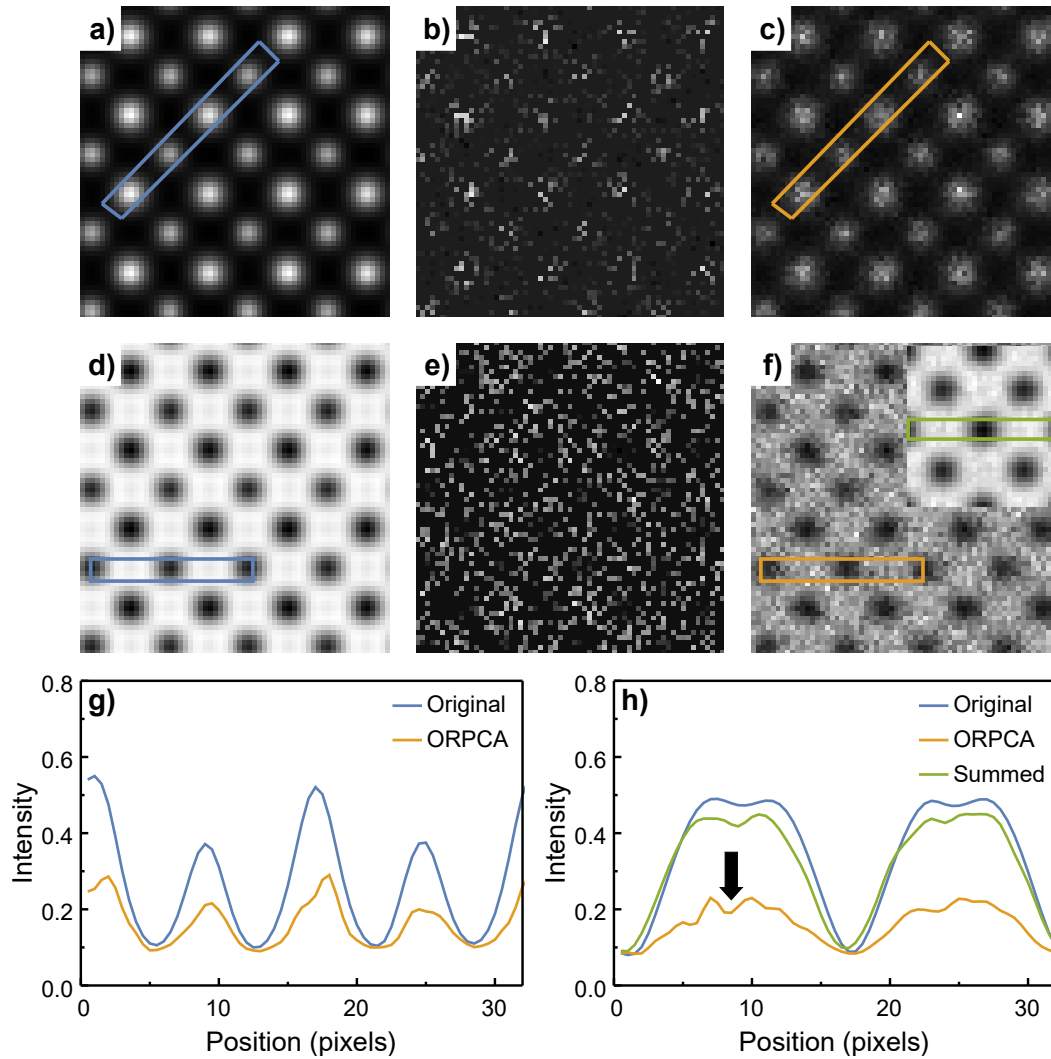
Although the data recorded directly from the camera can be compressed and stored efficiently (e.g. at low doses many pixels will be zero), most post-processing software packages and algorithms expect data as 32-bit or even 64-bit floats. This equates to approximately 20 GB of data requiring processing for every second of acquisition. Clearly, methods for denoising and dimensionality reduction that work in an online manner have a huge potential here.

Assuming the acquired dataset is indeed low-rank, which has been shown in this chapter to be a fair assumption for many electron microscopy datasets, we can make use of ORPCA to factor the dataset and store  $\mathbf{L}$  and  $\mathbf{R}$  only, while also denoising on-the-fly. For example, if  $r = 5$ , the memory requirement is reduced to 300 MB per second of acquisition, and individual frames can then be reconstructed when required. Although this memory reduction breaks down if there are major changes in  $\mathbf{L}$  over time (e.g. sample drift), the online denoising option remains an attractive capability. Note also that as discussed above, tensors are another option for reducing the storage requirements. Assuming the dataset has a low tensor rank, the flexibility of different tensor decomposition algorithms may bring even greater memory benefits than ORPCA [119].

#### 3.4.4 Inpainting and dose reduction

X-ray spikes, dead pixels and missing data are recurring problems in electron microscopy, and usual approaches include either manual identification and interpolation, or, in more extreme cases, the dataset is simply ignored and thrown away. The inclusion of the sparse term in ORPCA thus offers a robust, automated solution to dealing with data corruptions, and is therefore hugely promising for electron microscopy.

As well as unavoidable noise and corruptions, intentionally subsampling a dataset can bring experimental benefits. It was shown recently by Saghi et al. that by blanking the electron beam at random as the STEM probe scans, it is possible to recover a significant portion of the image using CS theory even when only a small percentage of pixels are sampled [171]. The motivation is that beam-sensitive specimens receive a much lower overall dose as each frame is acquired. Both the accumulated dose and the image acquisition time can be reduced by a factor of 5 or more using this approach [172]. Saghi et al. used a compressed sensing approach for filling in the missing pixels (also known as “inpainting”), operating only on individual images to recover the underlying features. However, if the low-rank constraints of robust PCA are satisfied then it offers an alternative solution, modelling the missing data as sparse errors, in a way that utilizes data correlations over time. This method of inpainting is similar to the matrix completion problem described by Candès et al. [135], but if time-resolved STEM imaging of beam-sensitive specimens is desired, ORPCA has the



**Figure 3.13:** Recovering a subsampled  $\text{SrTiO}_3$  image sequence with Poisson-Gaussian noise and sample drift. **(a)** A frame from the ADF sequence, **(b)** with Poisson-Gaussian noise added and 75% of the pixels removed, and **(c)** following recovery with ORPCA. **(d)** A frame from the ABF sequence, **(e)** with Poisson-Gaussian noise added and 75% of the pixels removed, and **(f)** following recovery with ORPCA. The inset is the result of an image sequence alignment and summation. **(g)** Line profiles from the highlighted regions of the ADF frames in **a** and **c**. **(h)** Line profiles from the highlighted regions of the ABF frames in **d** and **f**, with an arrow highlighting a dip in intensity. See also Movie S9.

capability to track and adapt to any changes over time such as adatom or defect motion, or sample drift.

Figure 3.13 depicts a frame from the same simulated  $\text{SrTiO}_3$  sequence as Figure 3.12, incorporating Poisson-Gaussian noise and sample drift. Also shown is the complementary ABF image (Fig. 3.13d). ABF imaging here reveals the light oxygen columns, which are not seen in the ADF image due to the dominance of Z-contrast by the heavier Sr and Ti–O atomic columns. Subsampling the image is achieved by masking with a matrix drawn from a Bernoulli distribution, with  $p$  pixels observed and  $1 - p$  pixels blank. For



this investigation only 25% of pixels were sampled, and the Poisson-Gaussian noise parameters were  $\alpha = 0.1$ ,  $\mu = 0$  and  $\sigma = 0.1$ . Despite the challenging noise and sample drift conditions, the ORPCA algorithm is able to successfully recover the main features from both the ADF and ABF sequences (see also Movie S9). Line profiles across the recovered images (Fig. 3.13g,h) confirm that while the overall image intensity is diminished, variations in contrast between elements are still present in the ADF case. Meanwhile, the ABF line profiles show that oxygen columns are discernible in individual recovered frames (see the dip at in the orange line marked by the arrow in Fig. 3.13h), and the aligned and summed image (Fig. 3.13f inset) shows a very clear dip in intensity over an oxygen column comparable to the original image. This last point is key, since performing image alignment directly on the subsampled image is much more difficult than the recovered low-rank image.

Note that the small fraction of observed pixels in Figure 3.13 actually means that the sparsity assumption behind RPCA does not strictly hold (the error component  $\mathbf{S}$  is not truly sparse). However, the low-rank property of the underlying data is strong enough such that the ORPCA algorithm is still able to recover the signal, again highlighting the flexibility of low-rank methods.

### 3.4.5 Foreground–background separation

A major application of RPCA in the field of computer vision is the separation of a video into the foreground, typically containing moving objects of interest, and the background [173]. A static or slowly-varying background is well-modelled as a low-rank matrix when each video frame is flattened into an image vector. Any moving objects of interest, on the other hand, are likely to constitute only a small fraction of the total pixels in each frame, and thus can be modelled as a sparse matrix. The ORPCA algorithm is therefore an interesting method for motion detection and object recognition systems, and especially for identifying and tracking the motion of objects at the nanoscale with electron microscopy.

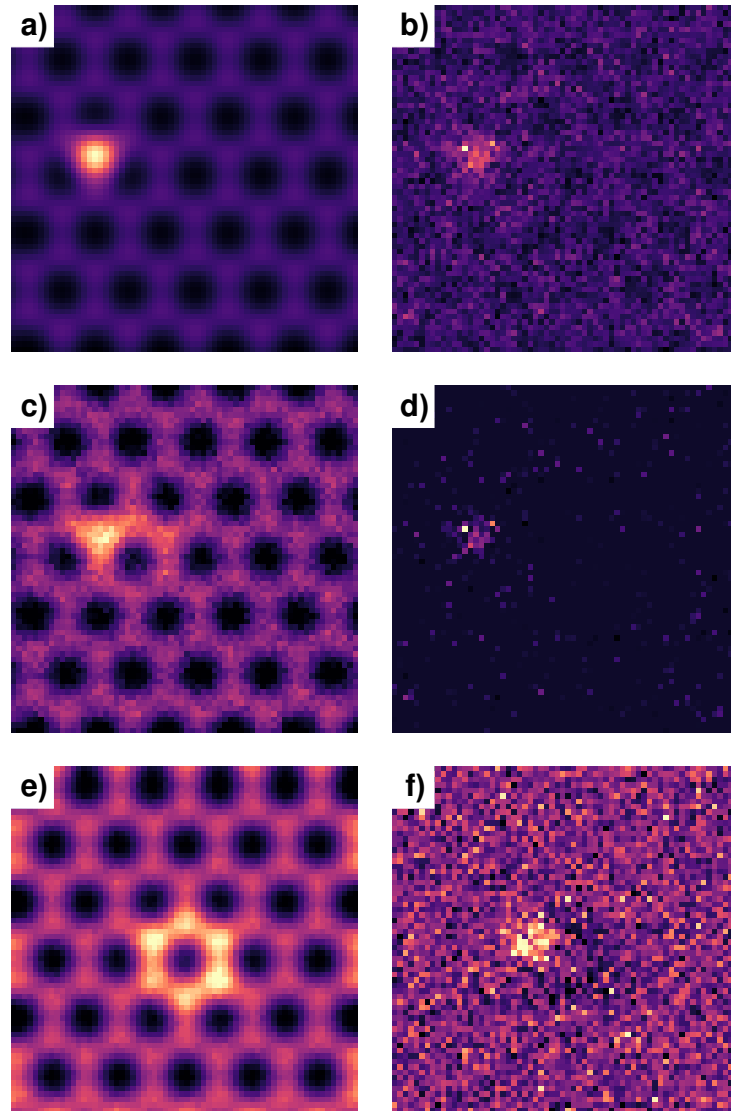
Inspired by the recent observation of a beam-driven random walk of a substitutional silicon atom in graphene [108], the ORPCA algorithm is applied here to separate the motion of an adatom from the background in the presence of noise and sample drift. Figure 3.14a shows a frame from a simulated, aberration-corrected ADF-STEM sequence exhibiting the same behaviour as Ref. 108. The full sequences is available as Movie S10. As in the previous  $\text{SrTiO}_3$  examples, sample drift was added to the sequence.

In Figure 3.14b, Poisson-Gaussian noise has been added to simulate the low-dose, short dwell-time imaging conditions. The second row (Fig. 3.14c,d) shows the low-rank component and sparse components estimated by the ORPCA algorithm combined

with the SGD method, with parameters ( $\lambda = 0.1$ ,  $\xi = 0.1$ , and learning rate = 1.1). The algorithm has recovered both components well, with reduced noise and a clear separation between the mobile Si atom and the underlying graphene substrate. The constant background level of Figure 3.14d highlights the benefits of the sparsity constraint used for the foreground. Use of the stochastic gradient descent solver has also enabled the algorithm to track and account for the sample drift, resulting in a clean image sequence suitable for subsequent alignment and tracking. By way of contrast, Figure 3.14e,f show the result obtained first by alignment and then summing the image sequence using an algorithm designed for STEM image alignment [174]. While the graphene lattice is easily recovered, the Si atom is harder to identify above the background noise (Fig. 3.14f).

### 3.5 Summary

This chapter has summarized the potential applications of low-rank machine learning methods to electron microscopy datasets. In particular, two novel algorithms have been demonstrated for the robust, automated denoising of microscopy datasets with a particular focus on time-resolved imaging and real-time processing in the presence of noise. The PGURE-SVT algorithm for denoising image sequences is applied to the study of adatom dynamics in Chapter 5 and again in Chapter 7, where the ORPCA algorithm is also employed to separate out adatom dynamics from the changes in the crystalline substrate underneath. Although the focus has been on time-resolved imaging, many of the techniques are equally applicable to other low-rank datasets. The noise estimation and parameter selection methods are even more general and can be used even if the underlying dataset is not low-rank. The key message is that a full understanding of the statistics of the acquisition process is central to the development of robust algorithms for processing electron microscopy data. This theme is continued in the next chapter, which develops methods for detecting, identifying and tracking atoms in time-resolved STEM imaging.



**Figure 3.14:** Separating the motion of a substitutional Si atom in graphene. **(a)** Sample frame from a simulated ADF-STEM sequence with sample drift. **(b)** The same frame with Poisson-Gaussian noise added ( $\alpha = 0.1$ ,  $\mu = 0$  and  $\sigma = 0.1$ ). **(c)** The low-rank component estimated by ORPCA. **(d)** The sparse component estimated by ORPCA. **(e)** The summed image after sequence alignment. **(f)** The difference when the summed image is subtracted from **b**. See also Movie S10.



## Chapter 4

# Tracking the motion of single atoms with ADF-STEM

### 4.1 Introduction

Bridging the gap between single atom electron microscopy and computational modelling requires not only the positions but also the identities of the atoms in the system to be known with the highest possible precision, in order for a direct comparison to be made between measured and calculated material properties. The addition of a temporal dimension also requires accurate tracking of atoms in order to analyze the dynamic behaviour of the system and link this with simulations. Accounting for the effects of noise, background and motion make the characterization and tracking a difficult task, and robust, automated methods are of particular importance for image sequences, where manual identification over many frames is prohibitive. Building on the theme of real-time analysis developed in the previous chapter, the ability to analyze the data in an online manner is an important consideration for in-situ experiments and those involving beam-sensitive specimens.

In this chapter, first methods for detecting atoms in image sequences are discussed, and a new classification approach based on image features is outlined. In the second part, a model-based approach for localizing atomic positions and calculating intensities is then presented. Particular attention is paid to the analysis of errors in positions and intensities, which are central to accurate system characterization. In the final part of the chapter, the important role of particle tracking in time-resolved single atom microscopy is highlighted, and a new online method for particle tracking is outlined with a focus on adaptability for different scenarios in STEM\*.

---

\*The atom classifier for detection (Sec. 4.2) and the tracking algorithm (Sec. 4.4) presented in this chapter were developed in collaboration with E. Schmidt, who wrote the Python code for the classifier and reinforcement learning methods.

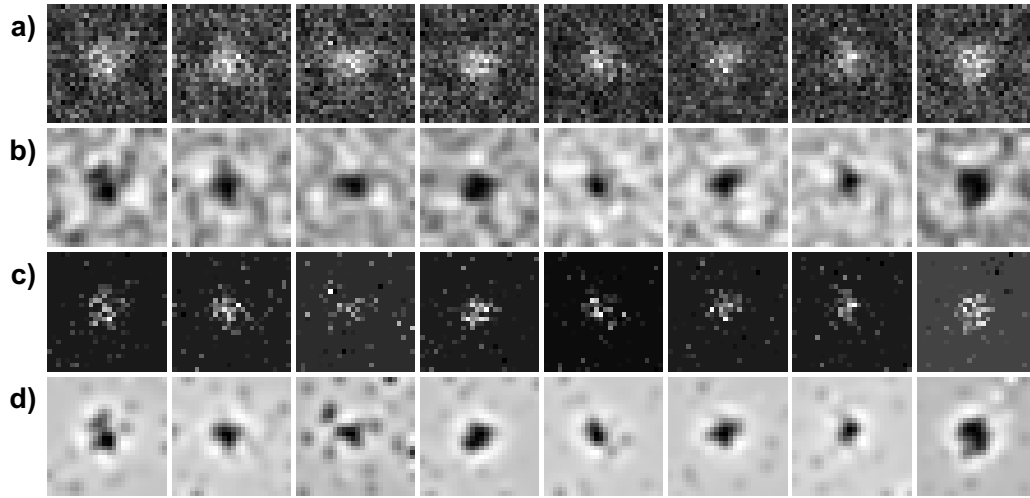
## 4.2 Robust detection of single atoms

After collecting one or more STEM images and applying any preprocessing such as denoising, the first step to atom characterization is to separate the pixels belonging to the objects of interest from the background. While the ORPCA algorithm of the previous chapter performs just this step, the low-rank and sparsity constraints mean performance is not always optimal, and other methods need to be considered. The incoherent imaging model of ADF-STEM makes the task somewhat simpler than in coherent modes such as BF or ABF imaging, since the absence of contrast reversal means the atoms appear, generally, as bright, round objects on a dark background. Atom identification in ADF-STEM therefore shares similarities with star detection in astronomical imaging and object detection in fluorescence imaging.

### 4.2.1 Detection approaches and challenges

In a noise, motion and aberration-free environment, with a constant background, identifying the atoms would involve a simple segmentation procedure according to a single intensity threshold. However, such a scenario rarely exists, and each experimental factor can introduce its own problems. Noise, as was seen in the previous chapter, can significantly obscure the objects of interest, particularly in low-dose scenarios. The noise reduction methods discussed in the previous chapter can be used to improve the image contrast, although this complicates any particle detection by introducing a dependency on additional parameters. A second problem, particularly relevant to time-resolved STEM but also to single images, is atomic motion, which, during the serial acquisition of a STEM image, can smear the intensity of an atom into broad, elliptical shapes. Fast scanning routines, low temperatures, and lower electron beam energies can all help to reduce this effect, although not necessarily eliminate it completely. Microscope aberrations can also lead to extended shapes of single atoms, although the development of aberration correctors has helped to reduce this problem.

The background level is particularly relevant to single atom imaging. Unlike the imaging of nanoparticles, where the atomic columns mean the intensity of the nanoparticle is typically many times higher than the background support, in single atom imaging the difference in  $Z$  between the foreground atoms and the background is often much smaller. The problems introduced by the background and noise typically rule out common detection approaches in fluorescence microscopy (where the background is zero) such as local maxima detection or Laplacian-of-Gaussian (LoG) filters, which highlight rapid changes in the second derivative of the image intensity such as edges. The background problem is highlighted in Figure 4.1, which applies a LoG filter to the noisy Si atom in graphene image sequence from the previous chapter. Applied



**Figure 4.1:** (a) A selection of cropped, noisy images from the simulated ADF-STEM image sequence of a substitutional Si atom in graphene (see Fig. 3.14), centered on the Si atom position. (b) Results of a Laplacian-of-Gaussian (LoG) filter applied to the images in a. (c) The same images following background subtraction with the ORPCA algorithm. (d) Results of a LoG filter applied to the images in c.

to the original image, noise is insufficiently suppressed and separating the intensity minimum (in the LoG-filtered image) attributed to the Si from spurious minima caused by noise is less than straightforward. Even when the ORPCA algorithm has been used to remove the background (Fig. 4.1c), there are numerous local minima obscuring the Si atom. Furthermore, the variations in shape, size and intensity make it difficult to determine an appropriate threshold to dictate what is or is not an atom, and this will be further complicated by the presence of two or more elements with differing ADF intensity.

While the robust PCA approach can be used to perform background subtraction prior to atom detection, it requires that the background is low-rank, and that the foreground atoms can be modelled within the appropriate sparsity framework. In scenarios where either constraint is violated, alternative detection methods are necessary. In the field of astronomical imaging, where objects are point sources convolved with the point spread function (PSF) of the telescope, a number of source extraction methods have been developed to separate the point sources from extended background sources, including the CLEAN and DAOPHOT algorithms [175, 176]. The Bayesian source extraction method of Savage & Oliver [177] uses the telescope PSF to determine the likelihood of each pixel belonging to either the background or to a point source, thus estimating the model and background in a single step.

Astronomical imaging algorithms are computationally attractive for large datasets with a large number of objects for detection. However, ADF-STEM atomic images differ from images of stars in that they are not strictly point sources convolved with a PSF. Instead,

they are formed by a convolution of the PSF with an object function that depends on the atoms (or atomic columns) being imaged [42]. Despite this distinction, the Bayesian source extraction method is often practical for aberration-corrected ADF-STEM images, since the object function is typically a sharp peak ca.  $0.3 \text{ \AA}$  in diameter, compared to a corrected probe  $1 \text{ \AA}$  in diameter [102]. Aberration correction means that the major contribution to the spatial resolution and atom shape is the broadening of the probe caused by the finite effective source size [65]. This is typically modelled as a Gaussian or Lorentzian function, although there are arguments that for a close match between experiment and simulation a linear combination of the two may be more appropriate [178].

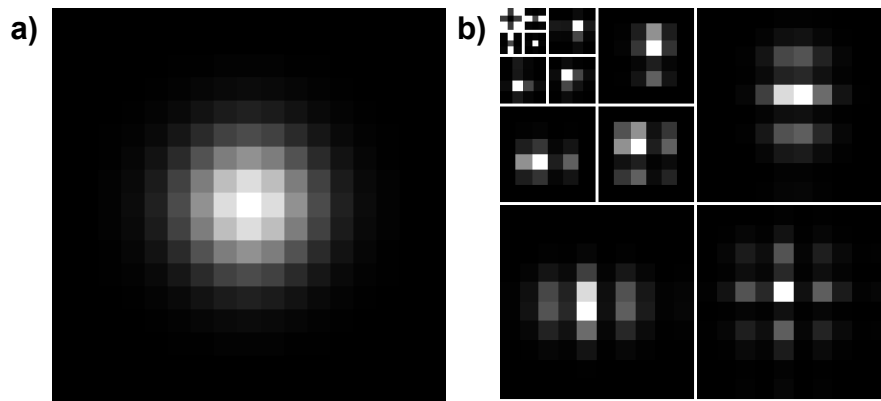
Searching the image for Gaussian or Lorentzian objects using astronomical methods is not always successful however. The serial acquisition of STEM (in contrast to parallel acquisition in astronomy) leads to an increased likelihood of the object of interest moving *during* a single frame, and denoising image sequences using the PGURE-SVT can, while improving the overall contrast of the image, further affect the shape of atoms. Departures from the symmetry of the idealized PSF are difficult to predict and will vary from atom-to-atom and from frame-to-frame, again leading to the problem of selecting appropriate thresholding parameters.

#### 4.2.2 A classification approach

It is clear that a “one size fits all” approach to atom detection would be inappropriate, while manual adjustment of parameters for each atom in each frame quickly becomes intractable as datasets increase in size. Taking inspiration from the wider fields of computer vision and machine learning, an alternative approach is to train an algorithm to recognize atoms on a small set of training data and subsequently apply it to larger datasets. This training step can incorporate prior knowledge such as microscope parameters and expected atom intensities, and could even use simulated STEM images to improve the recognition accuracy. This approach places the problem in a supervised learning framework (in contrast to unsupervised learning methods such as PCA), whereby a user indicates what the expected outcome should be, and the algorithm learns the relevant features and their relative importance in order to achieve this outcome [179, 180]. In this case the task is to classify image pixels as belonging to an atom of interest or the background, i.e. a binary classification problem.

Many methods have been developed over the years for tackling the supervised learning problem. For very large datasets containing millions of images or more, “deep learning” techniques based on neural networks have been shown to be extremely powerful tools, especially with the progress made in computer hardware with graphical processing units (GPUs) [181]. Neural networks reduce the necessary human input to





**Figure 4.2:** (a) A pixellated Gaussian representing an atom in an ADF-STEM image. (b) Applying a discrete wavelet transform using a Daubechies wavelet gives pyramid plot representing features in the image at different scales.

an appropriately labelled training set (which can easily extend to millions of images itself) and the network topology; the network learns what the relevant features are independently.

However, in the case of time-resolved electron microscopy, the available information is more limited, perhaps only a few hundred or thousand frames before beam damage and other experimental issues set in. While this might not be sufficient to train a large neural network accurately, it does make the input of an expert to guide the algorithm in the selecting the appropriate features much more practical, as the training set can simply be the first few frames of the sequence. The user labels a few samples of the image as being an atom (positive) or background (negative), and the algorithm utilizes this information to classify the remaining pixels in each frame. Standard object detection methods can be employed to build up a training set of positive and negative samples, but without the need for exhaustive parameter tuning on every image. The user next identifies the false positives and false negatives from the initial detections, and the labelled samples are used to build a larger training set by small shifts and rotations of the windows. This improves the invariance to object orientation and position within a window.

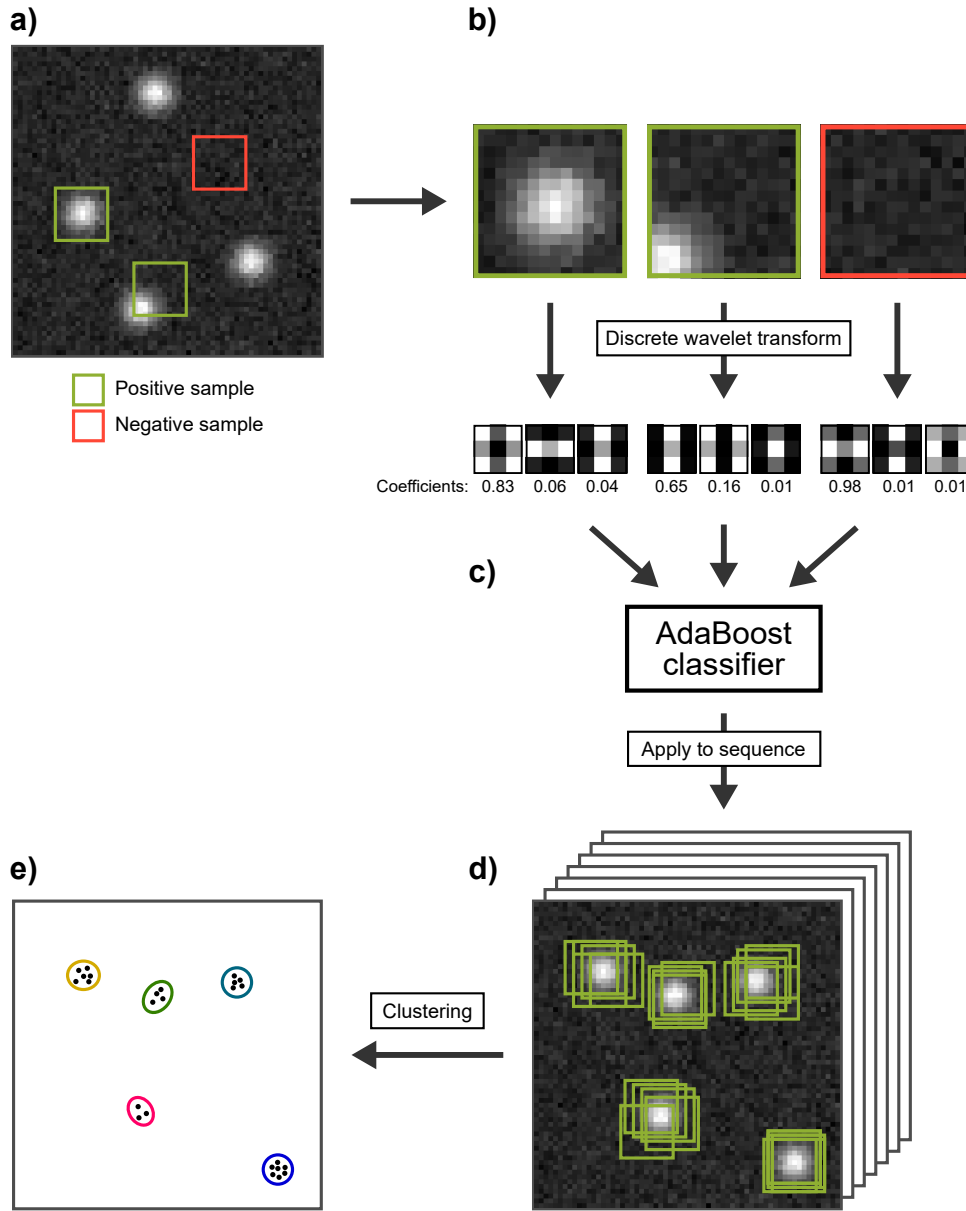
Further robustness to orientation, size and position can be achieved through a number of mathematical operations applied to the raw image. A useful method is based on the discrete wavelet transform, which combines spatial information from the pixel domain with frequency information from the Fourier domain to give a local description of the image features at different length scales. The simplest basis set for the transform is the Haar wavelet, but other wavelet families may be more appropriate for atoms by preserving roundness or some other property [182]; for example, Figure 4.2 applies the Daubechies wavelet family to a pixellated Gaussian representing an atom in an ADF-STEM image [183]. In Figure 4.2b, the features at different scales (from low-

to high-resolution) are shown in a pyramid plot. Each image in the pyramid will have a contribution to the original image weighted by a coefficient, rather like the idea of components in PCA from the previous chapter, and the coefficients can be used to distinguish between different features. Crucially, using feature vectors, rather than matching raw pixels, improves the invariance of the algorithm to any changes in magnification, noise, intensity or shapes in the window.

Building on an effective method for face recognition developed by Viola & Jones [184], Figure 4.3 presents the framework for identifying atoms in an image. In the training stage, initial small windows of the image are selected as positive and negative samples (Fig. 4.3a). Each sample window is then transformed into a feature vector using a wavelet transform, containing key information such as the total intensity and the presence of any edges or features that might distinguish an object of interest (Fig. 4.3b). The feature vectors along with their coefficients (or weights) are then passed to an algorithm known as AdaBoost (Adaptive Boosting) [185], which uses subsets of the full feature set to select and train the overall classifier using the best features, as a weighted linear sum of weaker classifiers (Fig. 4.3c). Each of these weak classifiers may only be slightly better than a coin toss, but the strength of the AdaBoost approach comes from how the weights are combined in the training phase. Initially each of the  $N$  samples in the data is given a weight of  $N^{-1}$ , but after each iteration the dataset is reweighted such that the classifiers focus on the incorrectly-classified samples from previous iterations, thus developing a stronger algorithm overall.

Figure 4.3d shows how the classifier is then applied to each frame of the sequence independently (enabling real-time operation), by sliding a small window of pixels across the image sequentially. Each window is again transformed and fed to the classifier, which outputs the positions of positive windows. After training and running the classifier, the final step is to turn these position into an estimate of the position of the atom. This is achieved by a hierarchical clustering of window positions, with the centroids forming the atom position estimate (Fig. 4.3e). This clustering can be controlled to incorporate knowledge about minimum distances between atoms, thus deciding whether a group of positive windows is a single atom or in fact two or more neighbouring atoms.

The classification method presented here provides a robust, fast and adaptable solution to the problem of atom detection. The results of this procedure can be used as an input for accurate localization and characterization via model-fitting, and later for particle tracking through image sequences. Although not discussed here, the classifier can easily be extended beyond a binary system to detect and group atoms or particles of different shapes, sizes and intensities, provided a training set can be constructed.



**Figure 4.3:** Schematic diagram representing the key steps of the classification method for atom detection. **(a)** A training set of labelled samples is generated by the user, perhaps using other object detection methods or ADF-STEM simulations. **(b)** The samples are transformed with a discrete wavelet transform to give a set of features describing the image, along with a coefficients describing the weighted combination of features. **(c)** The AdaBoost classifier is trained on the feature vectors. **(d)** The classifier is applied to the unseen image sequence using a sliding window. The positions of positively-classified windows are stored. **(e)** Hierarchical clustering is used to cluster the window positions, with the atom positions given by the cluster centroids.

### 4.3 Localizing and characterizing single atoms

Following identification of the atoms of interest, the next step is to improve the precision of the position, known as localization, and estimate the intensity attributed to the atom for Z contrast identification. The typical approach involves fitting some form of model to the observed pixel counts, with the model parameters including the atomic coordinates and intensity. This section looks at the model-fitting approach for ADF-STEM imaging of single atoms, and in particular considers the effects of detector noise and other experimental variables on the precision of the fitted model. Note here that while atom identification is typically performed on denoised data to minimize the number of false positives, the fitting described below is always performed on the original, unprocessed data. This allows improved quantitative analysis, and further enables the statistical treatment using a mixed noise model.

#### 4.3.1 Atomic model fitting

The incoherent model for ADF-STEM imaging (Eq. 2.1) means that individual atoms can be treated as independent scatterers, provided they remain well-separated. The image is then a convolution of the atomic potential with the microscope PSF, and by fitting this convolution to a small window around each detected atom, the atom position can be determined with sub-pixel accuracy. A typical and straightforward model for fitting atoms in aberration-corrected STEM is a 2D Gaussian [186, 187], based on the finite source size leading to Gaussian (rather than Lorentzian) broadening. The choice of a Gaussian in the present work is for consistency with previous studies only, since both functions can be computed efficiently, and the error analysis presented below is equally applicable to both scenarios.

Using the Gaussian model, the intensity attributed to an atom  $i$  at a given position  $(x, y)$  in an image is:

$$m_i(x, y) = \frac{V_i}{2\pi w_i^2} \exp\left(-\frac{(x - \bar{x}_i)^2 + (y - \bar{y}_i)^2}{2w_i^2}\right) + B \quad (4.1)$$

where  $(\bar{x}_i, \bar{y}_i)$  is the atom position and  $V_i$  is the total expected number of electrons scattered by the atom  $i$  onto the ADF detector. The standard deviation (or width)  $w_i$  can be estimated as part of the model fitting procedure, or it can be estimated separately, for example using data from known scatterers and the finite source size.  $B$  is a constant background level, although it can be replaced by a more general function  $B(x, y)$ . A constant background is not always appropriate, particularly in ADF-STEM, but for simplicity it is the approach taken here.

Although the model fitting approach is widely-used for  $Z$  contrast characterization, it can be very sensitive to the microscope and specimen parameters; whether an atom is modelled by a Gaussian or Lorentzian is one such example. An alternative approach is to use probe-integrated scattering cross-sections [188]. Here the dependence on microscope parameters is much smaller, although it does require a good estimate of atom positions (and indeed for all atoms to be correctly identified) for the segmentation step that assigns pixels to each atom to work correctly. The need for all atoms to be correctly identified means that the fitting approach is more appropriate here, and a straightforward method for dealing with missing or spurious atom detections is discussed below.

### 4.3.2 Maximum likelihood estimation

The next step is to develop a robust method for optimizing Equation 4.1 under the noise conditions encountered in electron microscopy. Previous approaches for atom-fitting have used a non-linear least-squares optimization algorithm to minimize the difference between the observations and the model [155]. However, such an approach is only guaranteed to give the optimal parameter estimates when the noise is Gaussian, and although weighted least-squares methods can mitigate some of the problems when the noise is non-Gaussian [189], the appropriate weights still need to be determined on a pixel-by-pixel basis. As discussed previously, a mixed Poisson-Gaussian noise model is a better way to characterize the noise in STEM, and the appropriate approach is to use maximum likelihood estimation (MLE) and directly incorporate the Poisson-Gaussian noise model, in a fashion similar to the unbiased estimators in Chapter 3. In the following discussion, an alternative definition of the detector gain is used for mathematical convenience compared to Equation 3.12, and the noise model is re-expressed as:

$$\mathbf{Y} = g\mathbf{Z} + \mathbf{E} \quad \text{with} \quad \begin{cases} \mathbf{Z} \sim \mathcal{P}(\mathbf{X}) \\ \mathbf{E} \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \quad (4.2)$$

where  $\mathbf{X}$  represents the original, noise-free image, and  $\mathcal{P}$  and  $\mathcal{N}$  correspond to the Poisson and normal distributions respectively. The gain  $g$  now represents the measured number of counts for every electron that arrives at the detector. For a noisy image  $\mathbf{Y}$  with  $K$  pixels, the likelihood of the observations being a result of the model  $m$  with parameters  $\theta = \{\bar{x}_i, \bar{y}_i, w_i, V_i, B\}$  is given by a convolution of a Poisson distribution and a normal distribution:

$$\mathcal{L}(\mathbf{Y}|\theta) = \prod_k^K \left( \frac{1}{\sqrt{2\pi\sigma^2}} \sum_q^\infty \frac{e^{-m} m^q}{q!} \exp \left( -\frac{\left( \frac{Y_k - \mu}{g} - q \right)^2}{2 \frac{\sigma^2}{g^2}} \right) \right) \quad (4.3)$$

The infinite sum in Equation 4.3 means a simple closed-form solution is unobtainable, but Huang et al. [190] showed that a simplification can be made by adding a constant  $\sigma^2/g^2$  to the model, such that the convolution is now:

$$\frac{Y_k - \mu}{g} + \frac{\sigma^2}{g^2} \sim \mathcal{P}(m) * \mathcal{N}\left(\frac{\sigma^2}{g^2}, \frac{\sigma^2}{g^2}\right) \quad (4.4)$$

Since the normal distribution in Equation 4.4 will asymptotically approach a Poisson distribution, using the relation  $\mathcal{P}(A) * \mathcal{P}(B) = \mathcal{P}(A + B)$  leads to:

$$\mathcal{L}(\mathbf{C}|\theta) \approx \prod_k^K \frac{\left(m + \frac{\sigma^2}{g^2}\right)^{C_k}}{\Gamma(C_k + 1)} \exp\left(-\left(m + \frac{\sigma^2}{g^2}\right)\right) \quad (4.5)$$

where  $\mathbf{C} = (\mathbf{Y} - \mu)/g + \sigma^2/g^2$  and  $\Gamma(x)$  is the standard gamma function. The optimal parameter estimates  $\hat{\theta}$  are then found by maximizing Equation 4.5. It is more common to use the log-likelihood instead, which replaces the product with a sum and also avoids numerical underflow when working with small likelihoods in floating-point arithmetic [191]:

$$\hat{\theta} = \arg \max_{\theta} (\ln \mathcal{L}(\mathbf{C}|\theta)) \quad (4.6)$$

This maximization can be carried out using a selection of non-linear optimization algorithms. However, it is important to be aware that the inclusion of the background term,  $B$ , or even a background function,  $B(x, y)$ , greatly complicates the analytical gradients of Equation 4.6. The alternative, numerical differentiation, amplifies the effects of noise and can thus degrade the performance of gradient-based optimizers [192]. Although knowledge of gradients typically gives better solutions than algorithms based on function values alone, many modern derivative-free methods are capable of very good solutions even in the presence of significant noise [193].

### 4.3.3 Estimating uncertainties in the model

Using the MLE approach facilitates the investigation of uncertainties in the model fitting, thus providing a useful assessment of the effect that noise and other variables have on the accuracy and precision of the atomic position estimates. This analysis is made possible by the Cramér-Rao Lower Bound (CRLB), which states that an unbiased estimator achieving the CRLB will give the lowest mean-squared error of all estimators for all possible values of  $\theta$  [194]. Furthermore, the variance of such an unbiased estimator is bounded by the inverse of the Fisher information matrix, which measures the amount of information about  $\theta$  that is carried by the observations [195]. The

covariance matrix  $\mathbf{K}$  is given by:

$$\mathbf{K}(\hat{\theta}) \geq \mathbf{F}(\hat{\theta})^{-1} \quad (4.7)$$

where  $\hat{\theta}$  is the solution to Equation 4.6.  $\mathbf{F}$  is the Fisher information matrix, with elements given by:

$$F_{ij} = -\mathbb{E} \left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln \mathcal{L}(\mathbf{Y}|\theta) \right\} \quad (4.8)$$

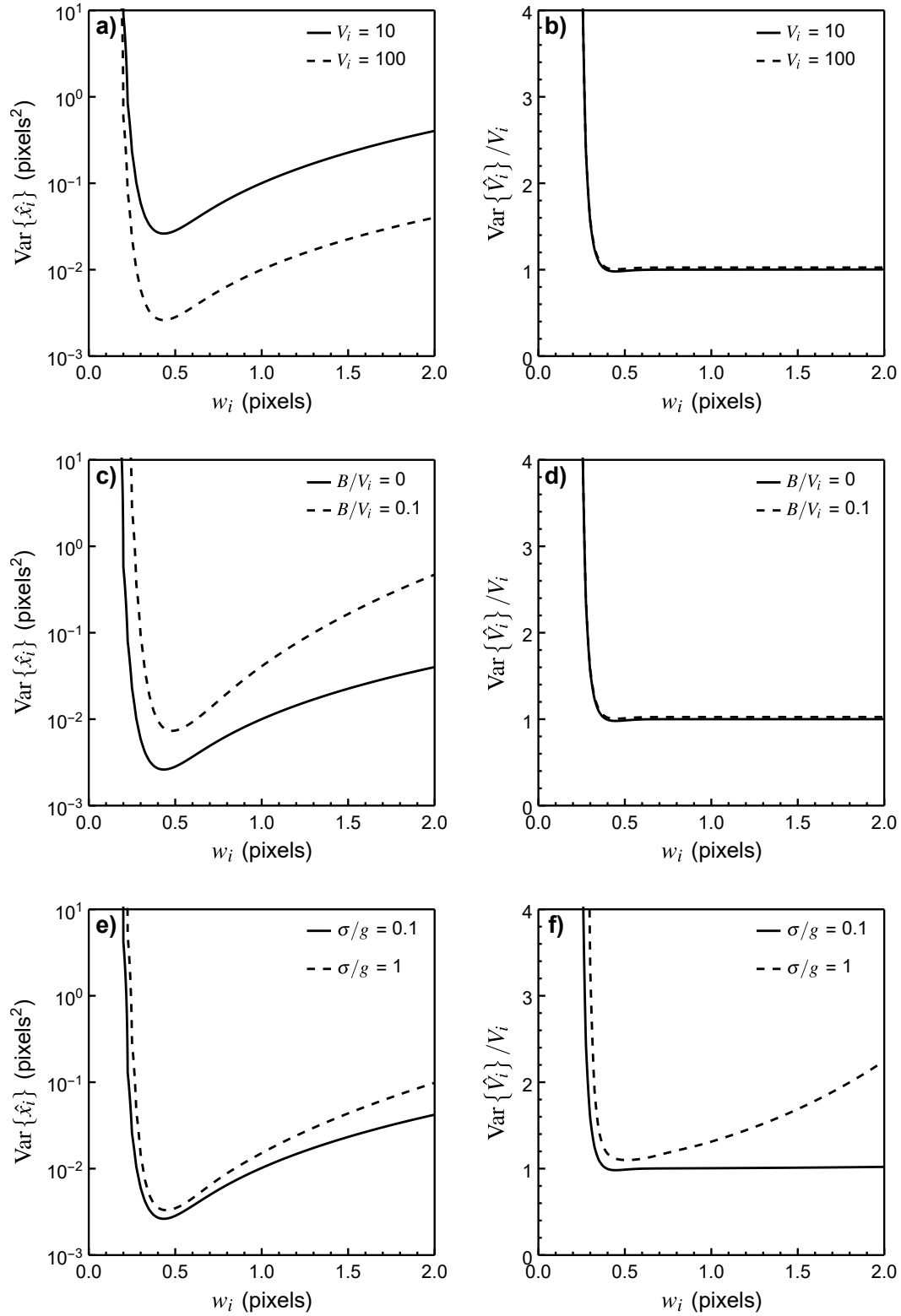
where  $\mathbb{E}$  indicates the expected value. In the case of Poisson-Gaussian noise, each element of  $\mathbf{F}$  can be approximated by the following expression [196]:

$$F_{ij} \approx \sum_k^K \frac{1}{m + \frac{\sigma^2}{g^2}} \frac{\partial m}{\partial \theta_i} \frac{\partial m}{\partial \theta_j} \quad (4.9)$$

If  $\sigma \ll g$  then this reduces to the Poisson noise case [197]. Analytical forms of the elements of  $\mathbf{K}$  are not derived here; instead, numerical solutions are used to explore the behaviour in Figure 4.4.

From Equation 4.1, the variance in the position  $\hat{x}_i$  can be plotted as a function of the atom width in pixels. Such a plot can then be used to determine the pixel size necessary to achieve a desired variance. The same procedure can be applied to another critical estimate for ADF-STEM imaging, the atom intensity,  $\hat{V}_i$ . In Figure 4.4a and b, increasing the number of counts in the model is seen to improve the precision of  $\hat{x}_i$ , but, as one would expect with Poisson noise, this has no effect on the *relative* variance of  $\hat{V}_i$ . The rapid increase in both plots as  $w_i \rightarrow 0$  is a result of the Gaussian intensity being concentrated inside a single pixel; when  $w_i = 0.25$  pixels, 90% of the counts are concentrated within a single pixel and localizing the centre of the Gaussian from essentially a single observation with any further accuracy is impossible.

The impact of the background  $B$  on the precision of  $\hat{x}_i$  (Fig. 4.4c) is not as significant as the effect it has on the intensity estimate (Fig. 4.4d). For a zero background level, increasing the number of pixels sampling the atom has no effect on the precision of the estimate  $\hat{V}_i$ , but this is no longer the case for the mixed noise model (Fig. 4.4f). The uncertainty introduced by the background increases significantly as the counts are spread over a larger number of pixels, suggesting that selecting an appropriate pixel size is key to good atom characterization using ADF-STEM Z-contrast. Previous work on single atom imaging argued that oversampling the imaged atom (i.e.  $w_i > 1$  pixel) allows it to be located with greater precision [51], while others argue the pixel size has no effect on the sensitivity to atomic number [198]. Figure 4.4 has shown this to be false, and oversampling can affect the quantification and characterization of the atom based on the ADF intensity. The second argument for oversampling is that atomic motion can be detected by shifts in intensity between adjacent scan rows. Clearly,



**Figure 4.4:** Measuring the behaviour of  $\text{Var}\{\hat{x}_i\}$  and  $\text{Var}\{\hat{V}_i\}$  as a function of the model width,  $w_i$ . **(a, b)** The effect of the total number of electron counts,  $V_i$ , with  $B = \sigma = 0$ . **(c, d)** The effect of the constant background,  $B$ , as a fraction of the total electron counts. **(e, f)** The effect of read-out noise,  $\sigma$ , as a fraction of the gain,  $g$ , with  $B = 0$ .



provided the atom can be identified and localized with sufficient precision, it is better to capture the movement from frame-to-frame rather than between scan rows if one is to robustly analyze the temporal behaviour. While localizing an atom with sub-picometre precision may be desirable for detecting cation shifts in ferroelectric materials (which are typically 10s of picometres), site-to-site atomic jumps are usually 1 Å or more, and reliably capturing the jump itself is arguably more important than the ultimate precision of the distance jumped. Figure 4.4 shows that oversampling needlessly exposes the sample to a much higher cumulative electron dose than is necessary to achieve a given precision, and validates the use of fast, coarse scanning routines for time-resolved ADF-STEM. Of course, increasing the pixel size (and correspondingly reducing the dwell time over a given area) results in an increase in noise, but this can be dealt with by the mixed noise model fitting methods described here.

#### 4.3.4 Fitting multiple atoms

For scenarios where the atoms are well-separated, it is sufficient to take a small window around each atom and fit the single model  $m_i(x, y)$ . However, when atoms are sufficiently close that the Gaussian tails start to overlap, a more appropriate approach is to model the electron counts at each pixel as a sum of  $N$  Gaussians (i.e.  $N$  is the number of atoms in the fitting region):

$$m(x, y) = B + \sum_i^N m_i(x, y) \quad (4.10)$$

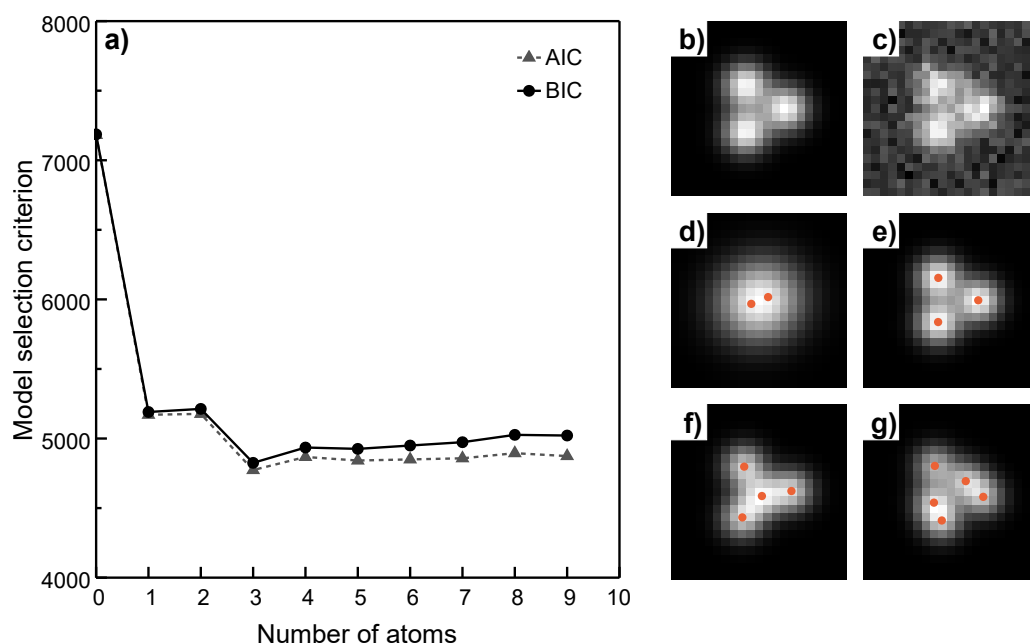
A crucial aspect of fitting multiple models to a region is ensuring that the model accurately represents the true number of atoms in the region, if  $N$  is unknown. Introducing more parameters to a model can lead to overfitting, where additional model parameters in fact explain the random noise in the observations. One way to avoid overfitting is apply a penalty to the number of model parameters. This can be achieved through hypothesis testing with log-likelihood ratios [179], or by using a measures such as the the Akaike Information Criterion (AIC) [199]:

$$\text{AIC} = -2 \ln \hat{\mathcal{L}} + 2n \quad (4.11)$$

or the Bayesian Information Criterion (BIC) [200]:

$$\text{BIC} = -2 \ln \hat{\mathcal{L}} + n \ln(K) \quad (4.12)$$

where  $\hat{\mathcal{L}}$  is the maximized likelihood from Equation 4.6,  $n$  is the number of model parameters ( $n = 4N + 1$  where  $N$  is the number of atoms if the background is



**Figure 4.5:** (a) The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) as a function of the number of fitted atoms in the model, where the true number is 3. (b) The original image. (c) The observed data following the application of Poisson-Gaussian noise. (d–g) Results from a two, three, four, and five atom model. The red dots indicate the peak locations found by the fitting algorithm.

constant), and  $K$  is the number of observations (i.e. the number of pixels in the fitting region).

Information criteria such as AIC and BIC thus utilize the maximum likelihood approach derived in the previous section to select the best model among a set of candidate models. Figure 4.5 shows how the criteria can be used to fit the correct number of atoms to an image. As the number of atoms,  $N$ , in the model increases, both AIC and BIC initially decrease (Fig. 4.5a). However, once  $N$  begins to exceed the true number of atoms in the image, the criteria are seen to increase and the optimization algorithm has trouble finding positions for these extra “atoms”. For example, in Figure 4.5f, a spurious peak is found between two existing peaks as a result of the algorithm attempting to fit the random noise. The BIC typically penalizes extra model parameters more strongly than the AIC, which is reflected in Figure 4.5a. Using model selection along with multiple model-fitting in this manner is a useful refinement for the particle detection approach, which separated positive sample detections into “atoms” using a clustering algorithm, with no information about the electron counts (i.e. the observables) incorporated. Atoms that lie within a certain distance of each other are prone to being classified as a single detection with the clustering method, so the model-fitting provides a secondary check for edge cases in a statistically robust manner.

### 4.3.5 Optimum pixel sampling

The 2D Gaussian model in Equation 4.1 describes a continuous function, whereas STEM data is acquired pixel-by-pixel in a discrete manner. In cases where  $w_i$  is small, this discrepancy is known to introduce a bias into the model [201], a problem that is reflected in the rapid increase in  $\text{Var}\{\hat{V}_i\}$  with decreasing pixel size as seen in Figure 4.4. Accounting for this bias, or identifying the point at which the effect becomes negligible, is achieved by first integrating Equation 4.1 across a pixel, so that the total number of detected electrons is equal in both the continuous and discrete models [202].

The adapted model for a square pixel of unit dimensions is given by:

$$m(x, y) = \frac{V_i}{4} E_{\bar{x}_i}(x) E_{\bar{y}_i}(y) + B \quad (4.13)$$

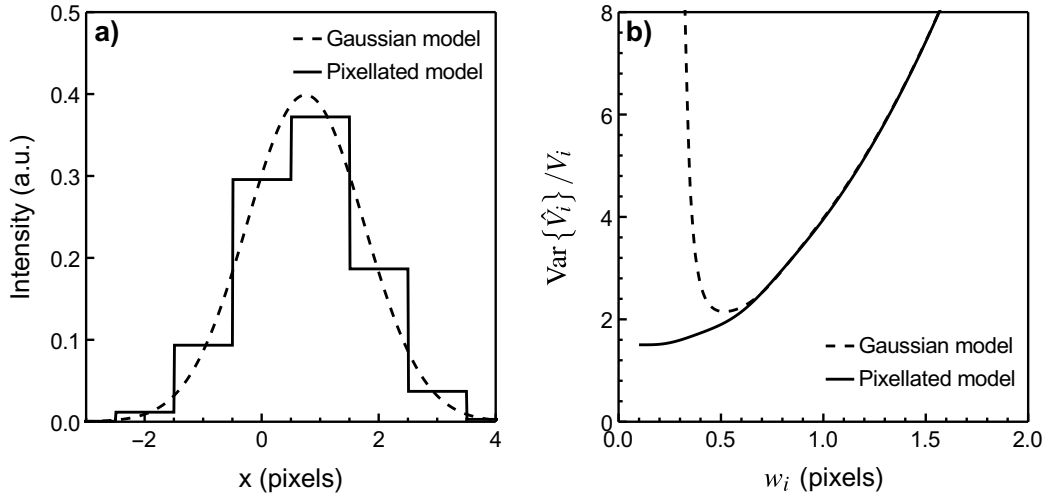
with the expression  $E_{\bar{x}_i}$  defined as:

$$E_{\bar{x}_i}(x) = \text{erf}\left(\frac{x - \bar{x}_i + \frac{1}{2}}{\sqrt{2}w_i}\right) - \text{erf}\left(\frac{x - \bar{x}_i - \frac{1}{2}}{\sqrt{2}w_i}\right) \quad (4.14)$$

$E_{\bar{y}_i}(y)$  can be defined in a similar fashion. Figure 4.6a compares a continuous Gaussian function with the pixellated model in one dimension. In Figure 4.6b, the advantage of the pixellated model for estimating the total electron counts  $V_i$  is evident when  $w_i < 1$  pixel. Practically, however, the problems caused by noise and the background are likely to dominate over the bias introduced by the pixel sampling, particularly as the atom first needs to be identified before it can be fitted. Equation 4.13 is thus perhaps more applicable to fluorescence microscopy than it is to ADF-STEM imaging, although as advanced sampling routines for low-dose STEM gain traction, analyses such as these are important to ensure experimental observations are interpreted correctly.

## 4.4 Automated tracking of atomic motion

Identifying and localizing the positions of single atoms only provides half of the picture needed for linking experiments with atomistic modelling. To fully unlock the temporal information obtained with rapid STEM imaging it is necessary to track the motion of atoms over time. The capability of time-resolved STEM was first explored by Isaacson et al. [13, 14], but even now atom tracking is typically restricted to just a few atoms over short timespans as a result of low SNR, beam damage and the lack of algorithms suitable for following large numbers of atoms.

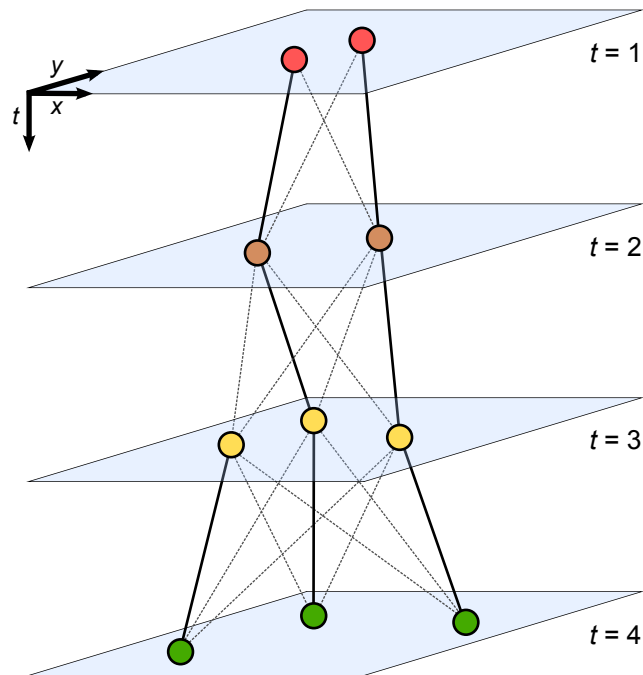


**Figure 4.6:** (a) Comparison between a continuous Gaussian model and the pixellated model (Eq. 4.14) in one dimension for  $\bar{x}_i = 0.75$  and  $w_i = 1$ . (b)  $\text{Var}\{\hat{V}_i\}$  as a function of  $w_i$  for the continuous and pixellated functions, with  $B = 0.1$ ,  $V_i = g = 1$  and  $\sigma = 0$ .

Recent developments in fluorescence microscopy, as well as other forms of bioimaging, mean that time-lapse imaging of intracellular processes can be performed with unprecedented spatial and temporal resolution. The size of these datasets, resulting from a large field-of-view and milli- or even microsecond frame rates, means that image sequences can contain thousands of particles, making manual tracking prohibitive and highlighting the need for robust, scalable computational methods. Similar problems also arise in astronomical imaging, where stars, galaxies and other objects are tracked over significantly larger length scales. These two fields have recognized the need for effective particle tracking algorithms, and the first part of this section briefly reviews some of the solutions in the literature. The second part focuses on the opportunities for online particle tracking, building on the work of Chapter 3 to investigate the possibility of designing a flexible workflow for denoising, detection, identification and tracking atomic-resolution STEM images in real-time at the microscope.

#### 4.4.1 Current methods for particle tracking

Successful particle tracking algorithms need to be able to cope with obscured particles, tracks crossing over each other and particles leaving or entering the field-of-view (Fig. 4.7), as well as account for particle motion models, which could be stochastic or directed in nature. The simplest method is the nearest-neighbour algorithm, which links together the closest particles between frames [203], while a more effective approach uses combinatorial optimization and “dummy” particles to incorporate missing or obscured particles [204].



**Figure 4.7:** The challenge of tracking particles through an image sequence involves not only identifying the correct linking between particles in consecutive frames (solid black lines) from the set of all possible links (dotted lines), but also dealing with scenarios such as new particles entering the field-of-view, as happens at  $t = 3$ .

In a recent review article comparing many different methods used in fluorescence microscopy [151], one of the best-performing algorithms took an approach based on multiple hypothesis testing (MHT). This applies a Bayesian framework to maximize the likelihood of the tracks given the set of observations [205]. Incorporating parameters such as motion models and the expected number of missing or new particles per frame enables a robust and general probabilistic framework for particle tracking to be constructed. The MHT approach is used for tracking the motion of copper atoms on graphene oxide in Chapter 5, as implemented in the Icy image analysis software package [206].

#### 4.4.2 Towards online tracking

Although the MHT approach performs well in many scenarios, the optimum method will depend on the noise conditions, particle motion dynamics, number of particles and background effects. This suggests that a method incorporating prior knowledge of the experimental system is likely to perform better than a more general algorithm. As well as using prior information and constraints to improve a tracking algorithm, there is also a strong case for an online approach, in a similar vein to the ORPCA

algorithm. An online approach not only means the algorithm can efficiently learn the system characteristics as the experiment develops, but also the instantaneous feedback for the user means the experiment can be guided based on the new knowledge. Most methods, including MHT, are designed to be used after the image sequence has been collected (although certain parameters are allowed to change through the dataset in an online manner). A particular problem with nearly all methods is that they assume either Brownian or directed motion of the particles. As will be discussed in depth in Chapter 5, these motion models are not always appropriate for the diffusion of adatoms on surfaces. A better approach would be to learn the appropriate motion model as the dataset is acquired, strengthening the case for an online algorithm.

#### 4.4.3 Reinforcement learning

The application of machine learning to object detection (Section 4.2.2) was a fairly typical binary classification problem: calculate the likely label (positive or negative) for a given input, and the algorithm compares this with the known labels of the training data to improve its performance. In many cases, however, the feedback may be non-specific, and the algorithm only knows if the classification is right or wrong but not *how* or *why*.

The approach for these non-specific scenarios is known as *reinforcement learning* [179], where the feedback is used to guide the exploration of the system. More formally, the computer learns the appropriate action to take in a given environment in order to maximize the reward (the non-specific feedback) [207]. The generality of this problem statement makes reinforcement learning widely applicable, provided the actions and reward can be specified. A famous example is in Google DeepMind's AlphaGo algorithm, which learnt how to defeat some of the best human Go players by playing many millions of games against itself, using a variety of strategies and moves, and adopting those that worked best [208]. The online strengths of reinforcement learning are best summarized by a trade-off between exploration of unknown regions of the environment with exploitation of current knowledge.

So how can reinforcement learning be applied to the particle tracking problem? The particle motion can be characterized with a joint distribution over space and time, which corresponds to the expected displacement and direction,  $\mathbf{r}$ , of a particle between consecutive frames, and the likelihood of a particle identified in one frame being the same particle in the second or third frame (thus incorporating missing or obscured particles). The joint distribution can be further extended to include intensity information for tracking atoms using ADF-STEM, since the  $Z$  contrast of an element should be similar between frames. Indeed, this facilitates tracking multiple classes of particles based on their ADF intensity.

The problem is then to find the set of links that maximize the likelihood of the experimental observations being drawn from this distribution. Taking two particles in different frames of a sequence, we define the action of linking them as  $\mathcal{C}_+$  and the action of not linking as  $\mathcal{C}_-$ . The simple, logical decision rule is to link the particles based on the ratio of the *prior* probabilities:

$$P(\mathcal{C}_+) > P(\mathcal{C}_-) \quad (4.15)$$

Of course, a decision rule based on knowledge about the system is always to be preferred over a rule based solely on prior probabilities. Let  $\mathbf{x}$  be a vector containing information about the displacement  $\mathbf{r}$ , the distance in time between the particles  $\Delta t$ , and any other information such as intensity. Then the likelihood of the observations  $\mathbf{x}$  depends on the underlying linking state,  $P(\mathbf{x}|\mathcal{C})$ . How does the measurement  $\mathbf{x}$  influence our knowledge about the true state of the system, i.e. whether the two particles are linked or not? Using new evidence to update the belief or knowledge of a system can be neatly expressed using Bayes' theorem [209]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4.16)$$

Applying Bayes' theorem, we now have for  $\mathcal{C}$ :

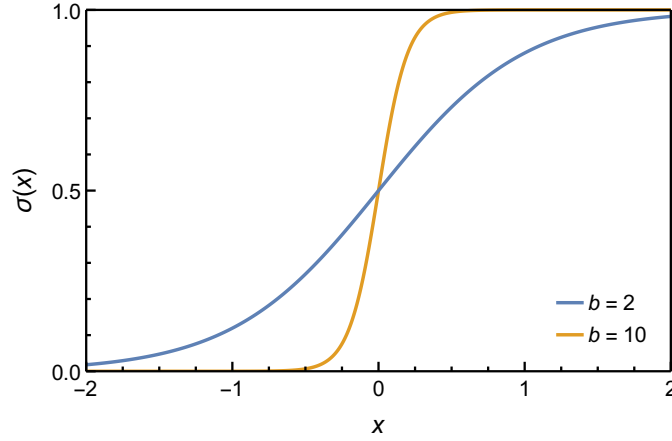
$$P(\mathcal{C}|\mathbf{x}) = \frac{P(\mathbf{x}|\mathcal{C})P(\mathcal{C})}{P(\mathbf{x})} \quad (4.17)$$

The expression  $P(\mathcal{C}_+|\mathbf{x})$  is known as the *posterior* probability, or the probability of the particles being linked given the observations  $\mathbf{x}$ . In a similar fashion,  $P(\mathcal{C}_-|\mathbf{x})$  gives the probability that the particles are not linked based on the available information. The denominator  $P(\mathbf{x})$  is the *evidence*, and is given by:

$$P(\mathbf{x}) = \sum_{\mathcal{C} \in \{\mathcal{C}_+, \mathcal{C}_-\}} P(\mathbf{x}|\mathcal{C})P(\mathcal{C}) \quad (4.18)$$

The prior probabilities  $P(\mathcal{C}_+)$  and  $P(\mathcal{C}_-)$  can be chosen based on how likely it is for a particle to actually appear again in another frame. The simplest case is a uniform prior,  $P(\mathcal{C}_+) = P(\mathcal{C}_-)$ . The argument for assigning an equal probability to linking and to not linking is that although it is uninformative, it will almost certainly perform better than setting  $P(\mathcal{C}_+) = 1$ , for example. By applying Bayes' theorem, our decision rule can now be updated to account for the observations:

$$P(\mathcal{C}_+|\mathbf{x}) > P(\mathcal{C}_-|\mathbf{x}) \quad (4.19)$$



**Figure 4.8:** The steepness of the sigmoid function (Eq. 4.20) is controlled by the parameter  $b$ , with the crossing point  $\sigma(0) = 0.5$ .

Choosing the correct linking for a particle in a given frame therefore requires the enumeration of Equation 4.19 for every possible link that can be formed and selecting the one that maximizes the future reward as expected in reinforcement learning.

The first step in the selection process converts the decision rule into a sigmoid function, which has the general form:

$$\sigma(x) = \frac{1}{1 + e^{-bx}} \quad (4.20)$$

The parameter  $b$  controls how steep the sigmoid function is while ensuring it always passes through (0,0.5) (Fig. 4.8). The purpose of the sigmoid function is to assign a “weight” to the possible link between the two particles based on the posterior probability, with the weight being used to select the link that maximizes the future reward. The cut-off value of  $\sigma(0) = 0.5$  means that the decision rule (Eq. 4.19) can be incorporated by replacing  $bx$  with\*:

$$\ln \left\{ \frac{P(\mathbf{x}|\mathcal{C}_+) P(\mathcal{C}_+)}{P(\mathbf{x}|\mathcal{C}_-) P(\mathcal{C}_-)} \right\} \quad (4.21)$$

Links with  $\sigma(\mathbf{x}) < 0.5$  are therefore automatically assigned a weight of zero based on the decision rule, since the probability of the particles being linked is less than probability of them not being linked.

Now a strategy is needed for selecting the correct link from the remaining possible links with a sigmoid value greater than 0.5. Remember that reinforcement learning seeks to maximize the *future* reward, which is related to the cumulative sum of the sigmoids. Naively selecting the new link as the one with the highest value of  $\sigma(\mathbf{x})$ , known as a greedy strategy, may maximize the current reward, but this action may ultimately

---

\*Notice that the decision rule (Eq. 4.19) can be written as  $\ln \left\{ \frac{P(\mathbf{x}|\mathcal{C}_+) P(\mathcal{C}_+)}{P(\mathbf{x}|\mathcal{C}_-) P(\mathcal{C}_-)} \right\} > 0$ .



reduce the reward at some later point. A simple alternative, known as  $\epsilon$ -greedy, chooses the maximum sigmoid link with probability  $1 - \epsilon$ , and otherwise chooses randomly from the set of non-zero weights with probability  $\epsilon$ . Introducing stochasticity in this fashion enables the algorithm to explore more of the environment than the greedy strategy, with the value of  $\epsilon$  controlling the amount of exploration. An alternative stochastic strategy known as *softmax* chooses a link with a probability, weighted by its sigmoid value, from a Boltzmann distribution [180].

After choosing a link between particles using the above strategy, the trajectories can be used to update the sigmoid by updating the likelihoods. In the case of  $P(\mathbf{x}|\mathcal{C}_+)$  this is a straightforward update using  $\mathbf{r}$ ,  $\Delta t$  from the current links, while for  $\mathcal{C}_-$  a sample of incorrect trajectories based on links that were not chosen can be drawn to update  $P(\mathbf{x}|\mathcal{C}_-)$ . This in turn updates the sigmoid  $\sigma(\mathbf{x})$ . Performing this update step every  $n$  frames rather than every frame allows the algorithm to learn the appropriate parameters for linking while using current and past knowledge. Choosing  $n$  then involves a consideration of the exploration–exploitation trade-off of reinforcement learning.

If the distribution or its parameters are unknown, then updating the sigmoid every few frames provides the flexibility required for learning the appropriate distribution over time, and the final learned sigmoid can be re-applied to the whole sequence at the end to refine the links formed early on. A small amount of manual tracking could also be used for training, without the significant cost of analyzing the whole dataset by hand. Alternatively, prior knowledge about the system is straightforward to incorporate into the algorithm, and not just in the priors  $P(\mathcal{C})$ . If the system is known to exhibit Brownian motion then the distribution of jumps between frames,  $P(\mathbf{r})$ , will be a Gaussian, whilst if the jumps take place on a regular lattice  $P(\mathbf{r})$  will be discrete. This information can be used to control the initial shapes of the sigmoid function.

A final, technical point is that the method can be optimized for real-time applications by using *kd*-trees for the linking, enabling an efficient search for nearby particles to consider for the update step [210]. This is justified by the fact that atoms at opposite sides of the frame or many frames apart in time are very unlikely to be linked, even when the spatial distribution is unknown, and so the use of nearby particles only acts as a loose constraint on the algorithm.

The reinforcement learning approach described here is intended as a general framework for online particle tracking. As with the classification algorithm, adaptability is key, and the use of a general sigmoid function to incorporate knowledge about the motion model makes it straightforward to apply any prior knowledge about the experiment.

## 4.5 Summary

This chapter has discussed some of the most important image processing and analysis challenges currently faced in single atom electron microscopy. Identifying atoms in noisy images is a complex task where no single algorithm may outperform others due to the extensive parameter tuning involved. Section 4.2 outlined an adaptable algorithm based on supervised machine learning, where much of the manual work is taken on by the algorithm. This leaves the user to focus on selecting the most appropriate image features for the task at hand, and facilitates the rapid processing of long image sequences in an automated manner.

Building on the atom detection algorithm, a method for robust localization and characterization of atoms using maximum likelihood estimation was presented. This showed that statistical analysis of the image acquisition process and noise characteristics can be used to guide the design of STEM experiments. Understanding what information needs to be collected means the microscope parameters can be optimized and the electron dose minimized, which is crucial for imaging beam-sensitive samples. Finally, the role of particle tracking algorithms in time-resolved microscopy was discussed. As with atom detection, no single algorithm is necessarily the “best”; it depends on the experimental conditions. By incorporating prior information about the system with reinforcement learning, an online algorithm for tracking was outlined, with built-in flexibility for different systems to ensure general applicability.

Above all, this chapter has, along with Chapter 3, highlighted the adaptability of machine learning techniques for different scenarios encountered in STEM, and how they can be combined with smart experimental design to push the boundaries of what electron microscopy can achieve. The next three chapters will apply these approaches to a series of experimental systems, where atom detection, fitting and tracking are key steps in the journey from experimental observations to a fundamental understanding of material properties at the atomic scale.

## Chapter 5

# Anomalous diffusion of copper atoms on graphene oxide

### 5.1 Introduction

Combining the capability for imaging and characterizing single atoms with the methods developed in the previous chapters makes STEM a powerful tool for investigating the dynamic behaviour of materials at the atomic scale. Dynamics are of interest over a wide range of timescales, but of particular importance and relevance to STEM imaging is the diffusion of atoms, both at surfaces and in bulk materials, where the typical frequency of atomic jumps is within the millisecond temporal resolution of time-resolved STEM. This chapter focuses on surface diffusion, which is fundamental to a wide variety of phenomena such as epitaxial growth, Ostwald ripening and heterogeneous catalysis. Recent interest in single-atom catalysts has made understanding surface diffusion at this scale particularly important, since it is a key process in the reactivity, stability and durability of such systems. While surface diffusion has been the subject of extensive experimental studies using FIM and STM [83], as well as theoretical modelling [211, 212], studies on industrially-relevant substrates are less common due to a number of experimental challenges including sample stability, robust identification of atoms, and limited temporal resolution.

In this chapter, low-voltage aberration-corrected STEM is used to excite and probe the motion of copper atoms on few-layer graphene oxide (GO)\*. In conjunction with experimental and computational characterization of the system, in-depth analysis of the dynamics reveals that the metal atoms undergo so-called anomalous diffusion, in contrast to the model of Brownian motion usually applied to the study of diffusive

---

\*The work presented in this chapter was carried out in collaboration with R.K. Leary, E.C. Tyo, S. Vajda, J.M. Thomas, Q.M. Ramasse, P.D. Bristowe and P.A. Midgley [213]. R.K.L. and Q.M.R. performed the ADF-STEM experiments. E.C.T. and S.V. prepared the copper samples.

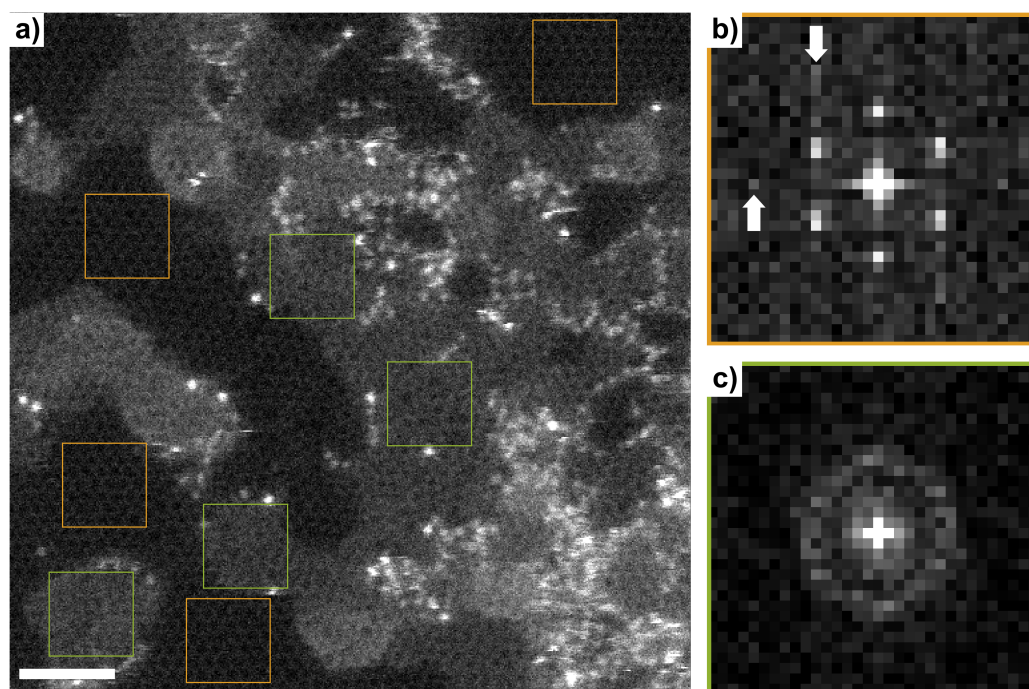
behaviour. This anomalous diffusion arises from the complex, heterogeneous nature of the GO substrate leading to a violation of the key assumptions behind Brownian motion. The observed behaviour is rationalized in the context of established stochastic frameworks in order to develop a model for the adatom–GO system, which is crucial to our understanding of diffusion of single atoms, and is a key first step towards controlling and harnessing dynamics at this length scale.

## 5.2 Atomic structure of graphene oxide

Graphene oxide (GO) is a form of functionalized graphene, the properties of which can be tuned by controlling the number and type of functional groups, opening up numerous electronic and optical applications [214–216]. GO is known to be spatially heterogeneous, with disordered and locally-varying oxygen coverage ranging from 10–50 at% [217], and there is great interest in metal–GO systems for heterogeneous catalysis, with oxygen-containing functional groups acting as preferential pinning sites [218]. For example, Fe atoms deposited on GO have been shown to be a low-cost and effective catalyst for the oxidation of CO [219]. However, a clearer picture of the substrate heterogeneity and the metal–substrate interaction is required to design the optimal catalyst in terms of reactivity and stability.

In this study, copper clusters were deposited on to few-layer GO membranes supported on standard holey carbon TEM grids (Agar Scientific), using a customized soft landing protocol [220]. Copper was chosen for two main reasons. Firstly, due to the expected strong interaction between the metal atoms and the oxygen atoms on GO, perhaps providing a route to mapping out the oxygen groups on the substrate. The second reason is that small Cu clusters are increasingly under investigation as a low-cost catalysts, in contrast to the noble metals commonly used in current heterogeneous catalysis systems [37, 221–223]. The copper atoms were nominally deposited as size-selected  $\text{Cu}_{13}$  clusters, but on imaging were found to be largely isolated single atoms. Chapter 6 will investigate the reasons behind this dispersion and any Cu–Cu interactions in greater detail, while the present chapter focuses on the interaction between single metal atoms and the GO substrate.

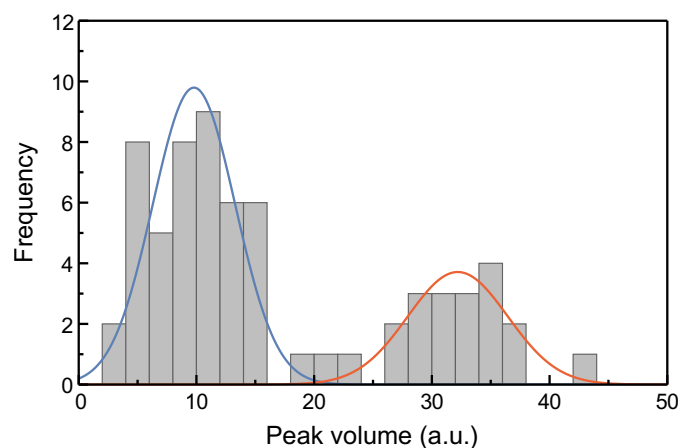
Figure 5.1a is an annular dark-field (ADF) aberration-corrected STEM image acquired at room temperature using a Nion UltraSTEM 100 operating at 60 kV. The illumination angle was 33 mrad and the probe current was 50 pA. The ADF inner detector angle was 85 mrad. The sample was baked for 8 hours at 120°C to drive off carbonaceous debris prior to observation. It is clear from this image that the GO is a complex material with a number of different environments available for copper adatoms. In agreement with previous (S)TEM studies [224–227], there are relatively pristine graphene regions,



**Figure 5.1:** (a) ADF-STEM image of copper adatoms on GO. Scale bar: 1 nm. (b) Summed Fourier transforms of the orange regions in a, with second-order Bragg reflections highlighted by arrows. (c) Summed Fourier transforms of the green regions in a.

with  $sp^2$  character (Fig. 5.1b), separated by brighter patches of poorly ordered material (Fig. 5.1c), likely to be oxygen-rich and thus exhibiting some  $sp^3$  character. Deposited adatoms are seen as bright dots in the image and are found almost universally at the boundaries of the disordered regions, suggesting very strongly that oxygen functional groups play a crucial role in the binding of adatoms to the substrate.

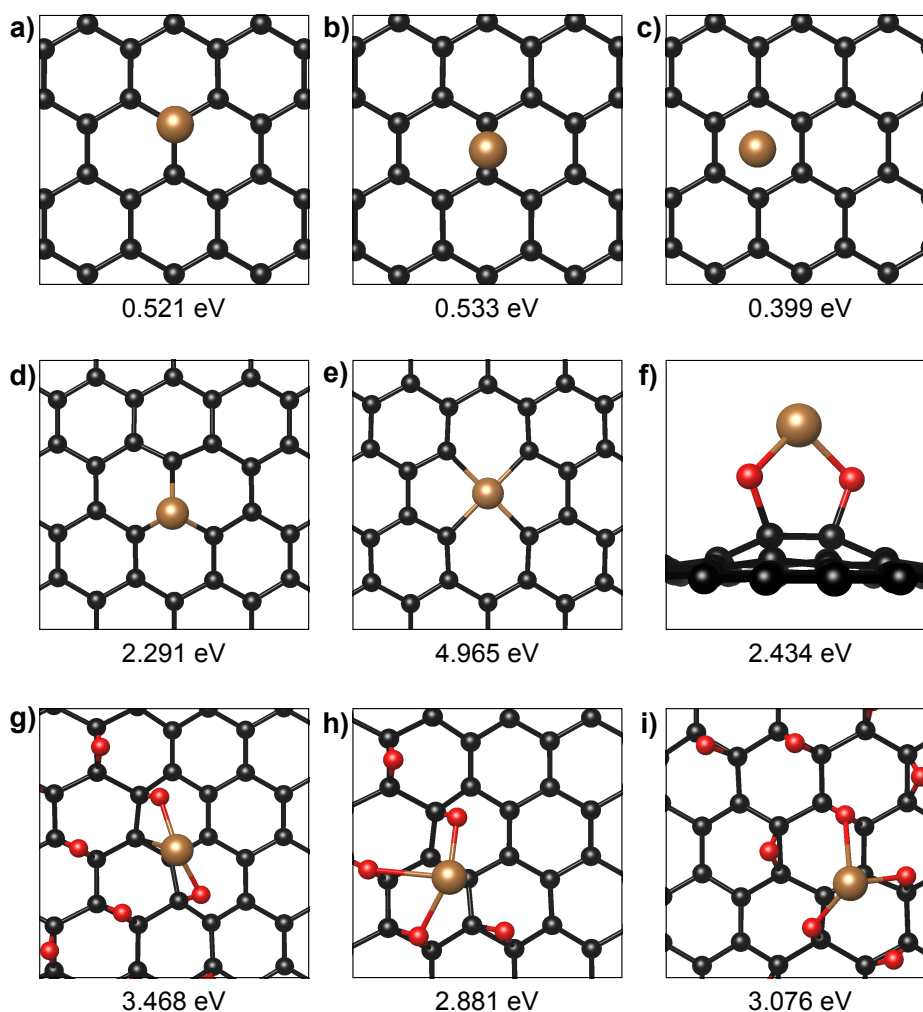
The high mobility of the adatom species under the electron beam made chemical identification using spectroscopy (either EELS or EDS) impractical due to the relatively long dwell times required for unambiguous fingerprinting. As discussed in Chapter 2, one of the advantages of ADF-STEM imaging is that intensity scales with atomic number  $Z$ , which, in favourable circumstances, enables chemical identification of atoms directly from image contrast at relatively short dwell times. Adatoms were identified in Figure 5.1, and a 2D Gaussian fitted to each atom as described in Chapter 4 assuming a Gaussian noise model (as the counts and therefore SNR in this image are relatively high). The distribution of Gaussian volumes has two distinct peaks, indicating the presence of two different adatom species (Fig. 5.2), and the ratio between the two peaks is 3.3:1. Taking the brighter atoms to be copper and the lighter atomic species to be silicon (a common impurity adatom on graphene oxide), an exponent of 1.65 gives a ratio of  $Z_{Cu}^{1.65} / Z_{Si}^{1.65} = 3.33$ .



**Figure 5.2:** Identifying adatoms using a histogram of ADF-STEM intensities. The solid blue and red lines are Gaussian curves fitted to the histogram.

To gain further insight into the nature of the binding between Cu adatoms and the underlying GO, a series of DFT calculations were performed. Further details on the DFT computations and the calculations of binding energies can be found in Appendix D. Figure 5.3 shows a number of possible low-energy configurations for a copper atom bound to both graphene and to GO. Copper is seen to interact weakly with pristine graphene (Fig. 5.3a-c), which is in line with studies of other transition metals on graphene [228, 229]. The energy barrier along the bridge-top-bridge pathway is only 0.012 eV, in contrast to  $k_B T = 0.026$  eV at room temperature. The adsorption is almost exclusively a result of van der Waals interactions, and the weak binding is reflected in the notable absence of Cu atoms on  $sp^2$  regions (Fig. 5.1); the atoms will diffuse rapidly at room temperature. The role of graphene vacancies in trapping Cu atoms was also explored (Fig. 5.3d,e), since vacancies are likely to form as a result of excitation by the electron beam, and are known to act as stable sites for other transition metal atoms [230].

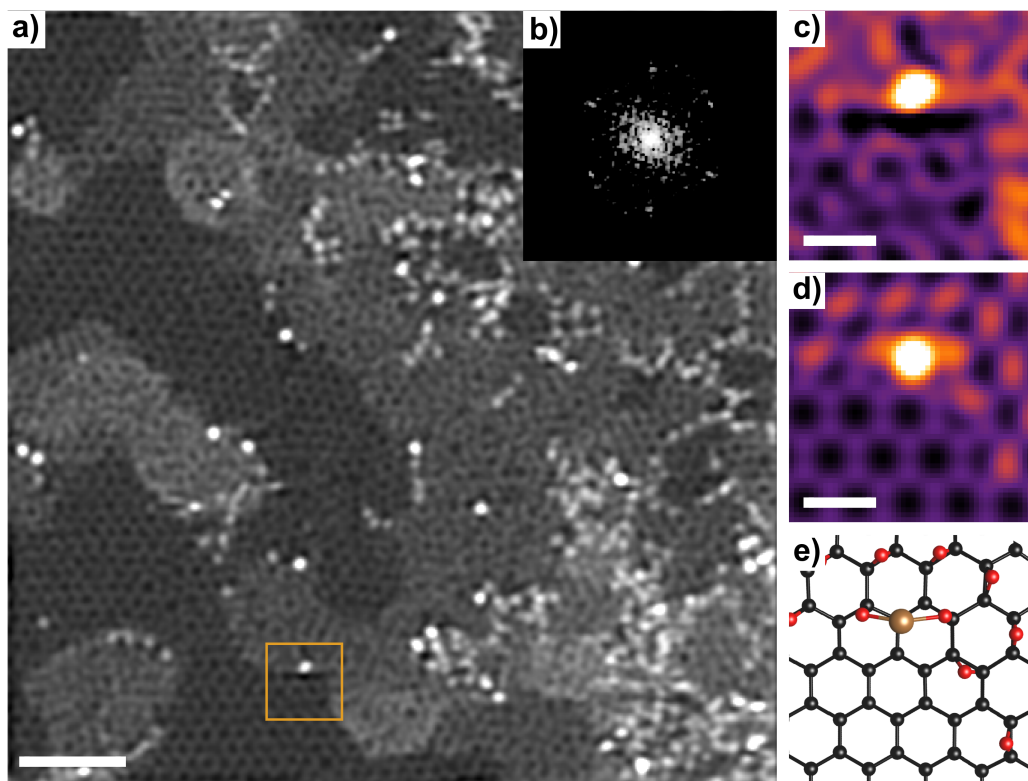
While vacancies and the high-symmetry adsorption sites of graphene are straightforward to model with DFT, the boundary regions of disordered GO (are favoured by adatoms in Figure 5.1) are harder to model due to the uncertainty over the positions of oxygen atoms on the substrate. As a result, the boundary regions were modelled using *ab initio* random structure searching (AIRSS) [231, 232]. A full description of the AIRSS technique is given in Section 6.4, but briefly, AIRSS generates a set of candidate structures by placing atoms at random positions, with physically meaningful constraints such as minimum interatomic distances. This is followed by local optimization with using DFT to find the low-energy structure. Applying this procedure led to a number of plausible pinning sites with binding energies of ca. 3 eV, some of which are shown in Figure 5.3f-i. These are similar in magnitude to the vacancy pinning sites, and importantly show that the binding of Cu to oxygen is much stronger than the



**Figure 5.3:** Density functional theory calculations of a copper atom interacting with graphene and GO. (a) Cu adsorbed in a top site. (b) Cu adsorbed in a bridge site. (c) Cu adsorbed in a hollow site. (d) Cu in a graphene monovacancy. (e) Cu in a graphene divacancy. (f) Cu bonded to two oxygen atoms on graphene. (g-i) Cu interacting with GO as calculated with AIRSS. The numbers are the binding energy of the Cu atom.

binding to graphene alone. The simulations presented here only considered oxygen on one side of the graphene sheet. While the presence of oxygen on both sides of the sheet will lead to a reduction in the buckling of the graphene [233], it is unlikely to result in significant changes to the binding between Cu and neighbouring oxygen atoms.

A further comparison between the experimental observations and the AIRSS simulations is given in Figure 5.4. By applying the double-Gaussian filter described in Ref. 50 to remove the contribution of probe tail effects, Figure 5.4a shows a clearer distinction between the ordered  $sp^2$  and disordered  $sp^3$  regions of GO. The local environment around a Cu adatom is enlarged in Figure 5.4c. One of the AIRSS structures, with bind-

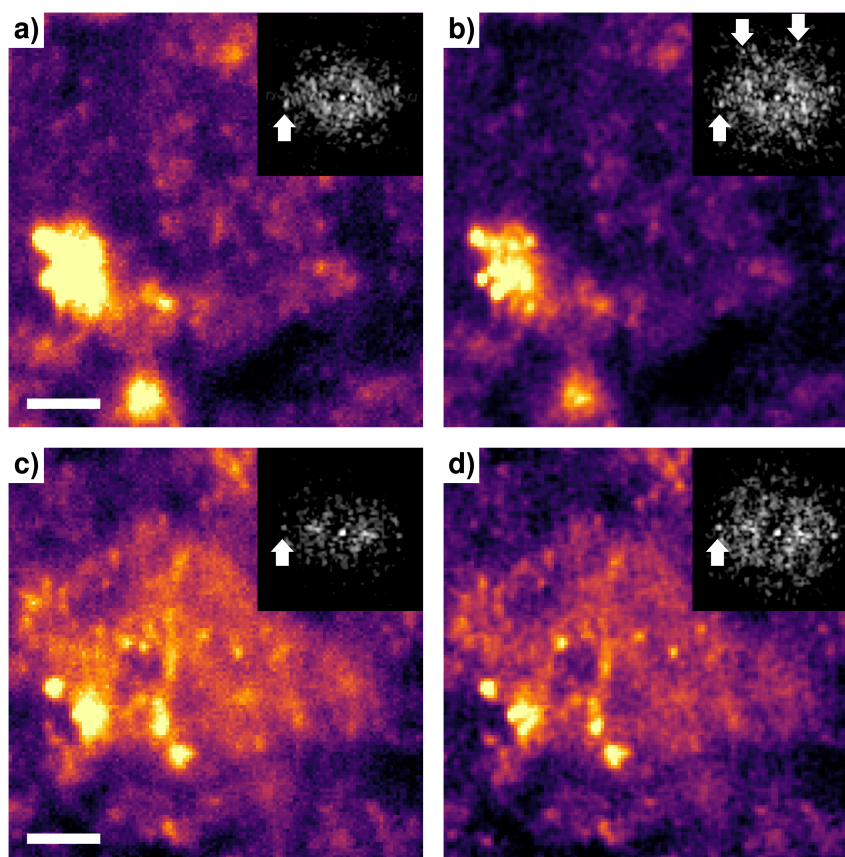


**Figure 5.4:** Heterogeneous structure of graphene oxide. **(a)** Filtered ADF-STEM image of adatoms on graphene oxide. Scale bar: 1 nm. **(b)** Fourier transform of the top-left quarter of the image, showing 6-fold symmetry. **(c)** Enlargement of the highlighted region, showing the local environment around a copper atom. Scale bar: 0.2 nm. **(d)** Simulated ADF-STEM image of a structure found with AIRSS, shown in **(e)**.

ing energy 2.85 eV (Fig. 5.4e), was used to simulate an ADF-STEM image (Fig. 5.4d), and shows remarkably good agreement with the experimental image (Fig. 5.4c).

This static ADF imaging has, in conjunction with DFT calculations, confirmed the complexity and heterogeneity of graphene oxide, and shown that describing the interaction of transition metal atoms with GO is not necessarily straightforward. Binding with oxygen functional groups and with graphene vacancies can pin adatoms to the surface, with differing impacts on the stability and possibly reactivity of the adatom species. Furthermore, even with modern “gentle” STEM techniques, it is very difficult to determine the exact local structure around a metal adatom with no ambiguity in terms of numbers and types of atomic species. The next section takes an alternative approach to understanding the system by recording the dynamics of adatoms as they explore their local environment.

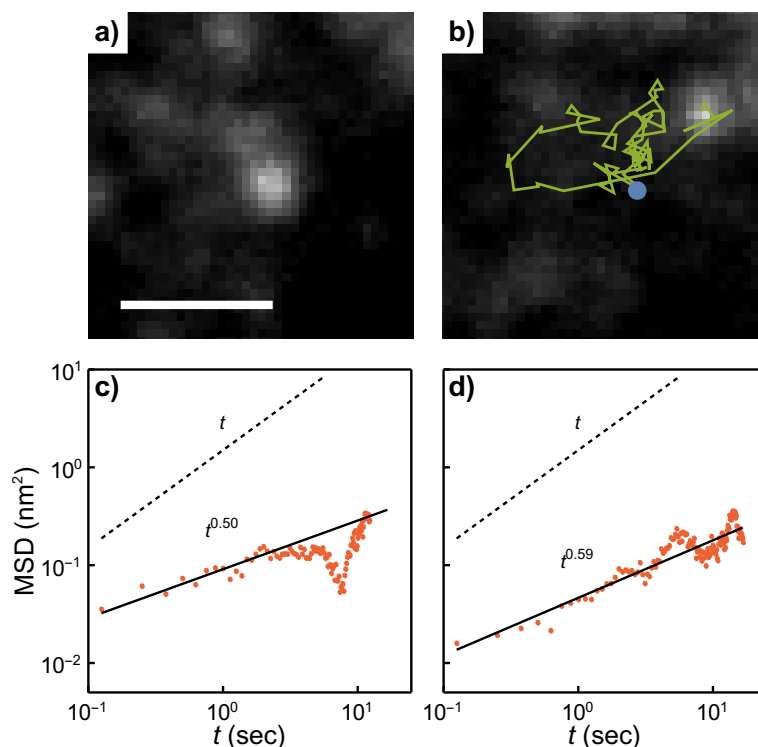




**Figure 5.5:** Evidence of graphene oxide in the image sequence. **(a)** A summation over all frames from sequence 1 before image alignment to the stationary atom. Brightness increased to highlight substrate structure. Scale bar: 1 nm. **(b)** Summed frames after alignment. **(c)** Summed frames from sequence 2 before alignment. **(d)** Summed frames after alignment. The insets are the Fourier transform of the respective images, with first-order Bragg reflections highlighted by arrows.

### 5.3 Capturing single atom dynamics with ADF-STEM

In conjunction with the state-of-the-art low-voltage STEM imaging in the previous section, further investigations were carried out at a primary electron energy of 80 keV to probe the dynamics of copper adatoms on the surface of GO. The maximum energy transferred in a collision between an accelerated electron and a nucleus is given by Equation 2.2, and for an 80 keV electron colliding with a copper atom,  $E_{\max} = 2.98$  eV. This is comparable to the binding energies calculated for various Cu pinning sites in Figure 5.3 and, acknowledging that the electron beam can also induce changes in the GO substrate, means that the beam can be used to excite the copper atoms out of pinning sites and subsequently record the movement. Accordingly, the subsequent atom motion may be considered as comparable to that likely to be seen at elevated temperatures. One must of course recognize that the conditions do not exactly reproduce those typical of catalysis; even in advanced environmental STEM the achievable gas pressures are a

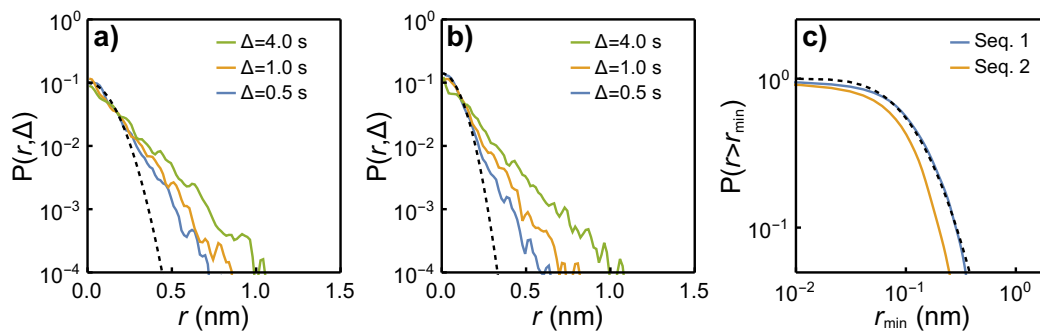


**Figure 5.6:** (a) Cropped region from the first frame of sequence 1 after denoising. Scale bar: 1 nm. (b) The same region 10 s later, with a copper atom trajectory highlighted in green. The blue circle highlights the starting position. (c,d) MSD of the trajectories extracted from sequence 1 and sequence 2 respectively. Best-fit lines are indicated in black with corresponding exponents, and the dashed lines show the behaviour expected from normal diffusion.

tiny fraction of atmospheric conditions. However, the idea of exciting and recording atom dynamics has been applied successfully to the study of numerous materials using aberration-corrected STEM, both in the case of adatoms on surfaces [107, 234], and for dopants in bulk materials [110].

Two ADF-STEM image sequences were recorded on an FEI Titan<sup>3</sup> from the same area of the sample with a 40 s wait between the two acquisitions. The probe convergence semiangle was 21 mrad, and the ADF inner angle was 36 mrad. Each sequence was recorded with a STEM probe size of ca. 0.2 nm, at 8 fps. The estimated dose was 2200  $e^- \text{ pixel}^{-1}$ , corresponding to a dose rate of  $6 \times 10^6 e^- \text{ nm}^{-2} \text{ s}^{-1}$ . Although each individual image in the sequence has a poor SNR, correlations between frames were used to reduce the noise via low-rank matrix recovery with the PGURE-SVT algorithm. The full sequences are available on the accompanying CD (see Movies S11 and S12).

Despite the PGURE-SVT processing, the SNR is still too low to discern the atomic arrangement of the GO substrate in the same manner as Figure 5.4, but its graphitic character, and underlying hexagonal symmetry, can be seen following alignment of the



**Figure 5.7:** (a) The distribution of jump lengths,  $r$ , at different intervals  $\Delta$  for sequence 1. The dashed line is a Gaussian fit to the data for  $\Delta = 0.5$  s. (b) The same analysis for sequence 2. (c) The cumulative distribution of jump lengths for both sequences, and a fit of a log-normal distribution to sequence 1 (dashed line).

sequences. Analysis of the sequences reveals that one of the adatoms remains pinned in the same site for the full duration of each sequence. By aligning the frames to the mean position of this atom, it is possible to account for the drift of the sample over time. Summation of the frames following this alignment reveals evidence of a graphitic lattice (Fig. 5.5). The summed images also reveal a change in the substrate as a result of electron beam irradiation between sequence 1 and sequence 2. The increase in brightness in the central region of sequence 2 (Fig. 5.5d) could be a result of increased oxide coverage or hydrocarbon contamination as a result of irradiation by the electron beam.

The positions of the copper atoms in each frame were extracted and linked to form atomic trajectories using the MHT method described in Chapter 4; the green track in Figure 5.6b shows how a single copper adatom has diffused over 10 s. In total, 63 trajectories were extracted from the two sequences with lengths of 25–130 frames, corresponding to ca. 3–16 s. Sample drift was removed from the atomic trajectories in a manner similar to the image alignment in Figure 5.5. A useful method for analyzing particle trajectories involves calculating the mean-square displacement (MSD) as a function of time. This can be used to assess whether the atomic motion is random, i.e. diffusive, or directed, perhaps as a result of being pushed by the electron beam. The MSD for the two sequences is plotted in Figures 5.6c and d. The standard model for diffusive motion predicts that the MSD is linear in time. It is clear from Figure 5.6 that the present system exhibits a significant departure from standard model, with exponents of  $0.50 \pm 0.11$  and  $0.59 \pm 0.04$  for sequences 1 and 2 respectively. Along with a recent TEM study of Fe atoms moving along graphene edges [235], this represents one of the first observations of so-called *anomalous* diffusion behaviour at the atomic scale.

Further evidence for anomalous diffusion is obtained by considering the distribution of atomic displacements, which in standard diffusion models are assumed to take

place either on a regular lattice or are drawn from a normal distribution. Analysis of the particle displacement distributions for various intervals for each sequence (Fig. 5.7a,b) indicates that this is not the case, with both heavy tails and hints of preferred distances due to the underlying GO substrate. Analyzing the cumulative distribution of jump lengths (Fig. 5.7c), i.e. the likelihood of displacements being larger than a certain value, shows that for both sequences the data closely follows a log-normal distribution, indicating that the jumps are uncorrelated processes with multiplicative probabilities [21]. This evidence is important for determining the underlying cause of the diffusive behaviour. The next section consider the theories behind both normal and anomalous diffusion, which are necessary for the development of a model to explain the experimental observations.

## 5.4 Analyzing and modelling atomic diffusion

The first descriptions of the random motion of microscopic particles were made by Robert Brown in 1828 [236], and the behaviour is now commonly known as Brownian motion. Brown recorded the motion of pollen grains suspended on the surface of water under an optical microscope. Despite no obvious mechanism for motion, the pollen grains were seen to dance about erratically. Further observations of dust particles indicated that the behaviour was not confined to living systems, but Brown was not able to ascertain the reasons for the motion. It wasn't until much later that Einstein used the random motion of large particles to infer the existence of atoms [237]. We now know that the motion Brown observed was caused by collisions between energetic water molecules and the much larger pollen grain, and the net effect of many millions of collisions is to push the grain in different, random directions. Einstein's analysis thus constitutes the start of a significant body of work on the stochastic behaviour of particles at microscopic scales. Observing and understanding the motion of single atoms continues these investigations to the smallest length scales, and modelling the diffusion of single atoms is key to understanding and predicting the material properties based on this motion.

It is also important to note that the phenomenon of Brownian motion, and indeed all of the diffusion models discussed in this chapter, can instead be framed in a more mathematical context as a random walk [238]. Typically Brownian motion refers to the natural phenomenon, while the mathematical model is known as a Wiener process. One of the first examples of modelling Brownian motion as a stochastic process was the study of stock option prices performed by Bachelier in 1900 [239]. This chapter focuses on the physical models of atomic diffusion on surfaces, although many of the terms and mathematical expressions are applicable to both frameworks. In-depth discussions of stochastic processes can be found in Refs. 240 and 241.

### 5.4.1 Normal diffusion

Einstein's analysis of the behaviour of a diffusing particle was based on the diffusion equation, which is equivalent to Fick's second law [237]. In two dimensions, let  $P(\mathbf{r}, t)$  be the probability of finding a particle at position  $\mathbf{r}$  at time  $t$ . The diffusion equation is then:

$$\frac{\partial}{\partial t} P(\mathbf{r}, t) = D \frac{\partial^2}{\partial \mathbf{r}^2} P(\mathbf{r}, t) \quad (5.1)$$

where  $D$  is the diffusion coefficient. If we assume  $\mathbf{r} = 0$  at  $t = 0$ , the solution to Equation 5.1 is the probability density function (PDF) of the normal distribution. The first moment, or mean, of this distribution is:

$$\langle \mathbf{r}(t) \rangle = \int_{-\infty}^{\infty} \mathbf{r} P(\mathbf{r}, t) d\mathbf{r} = 0 \quad (5.2)$$

A diffusing particle is therefore equally likely to move in all directions. The second moment, or variance, of this distribution gives the MSD of the particle at time  $t$ :

$$\langle \mathbf{r}^2(t) \rangle = \int_{-\infty}^{\infty} \mathbf{r}^2 P(\mathbf{r}, t) d\mathbf{r} = 4Dt \quad (5.3)$$

This is the linear relationship highlighted in Figures 5.6c and d. Einstein then derived the relationship between  $D$  and Avogadro's number,  $N_A$ , as  $D = \mu k_B T$ , where  $\mu$  is the atomic mobility,  $k_B$  is Boltzmann's constant and  $T$  is temperature\*. The first experimental measurements of  $N_A$  were made by Perrin in a series of microscopic tracking experiments that applied Einstein's analysis [243].

### 5.4.2 Anomalous diffusion

In deriving Equation 5.1 and the resulting properties above, several key assumptions were made about the system under investigation:

- particles are independent
- particle displacements are symmetric (cf. Eq. 5.2)
- particle displacements are statistically independent

These assumptions result in the normal distribution PDF being the solution to Equation 5.1, and Einstein's model for Brownian motion is otherwise known as *normal* diffusion. For the motion of pollen particles suspended in water as reported by Brown in 1828, it is straightforward to see the validity of these assumptions. The particles are typically well-dispersed, and are free to move in all directions. Finally, the observation

---

\*The same result was derived independently by Smoluchowski [242].

timescales are sufficiently large compared to the collision timescales such that each observed displacement can indeed be considered independent.

Violations or departures from any of the three assumptions above can give rise to a wide range of behaviours known as anomalous diffusion. As experimental techniques are developed to explore particle dynamics at ever-smaller length and time scales, it is no great surprise that the assumptions in the Einstein-Smoluchowski model begin to break down. For example, extensive work with single-particle tracking using fluorescence microscopy has revealed that the crowded, active environment inside cells means that diffusing molecules can no longer be treated as independent, displacements take place in preferred directions, and individual jumps are correlated [21, 244–246]. These violations result in anomalous, rather than normal, diffusion.

In contrast to the Einstein-Smoluchowski model, anomalous diffusion is characterized by a non-linear relationship between the MSD and time. Usually, anomalous diffusion takes the form of a power-law, so in two dimensions:

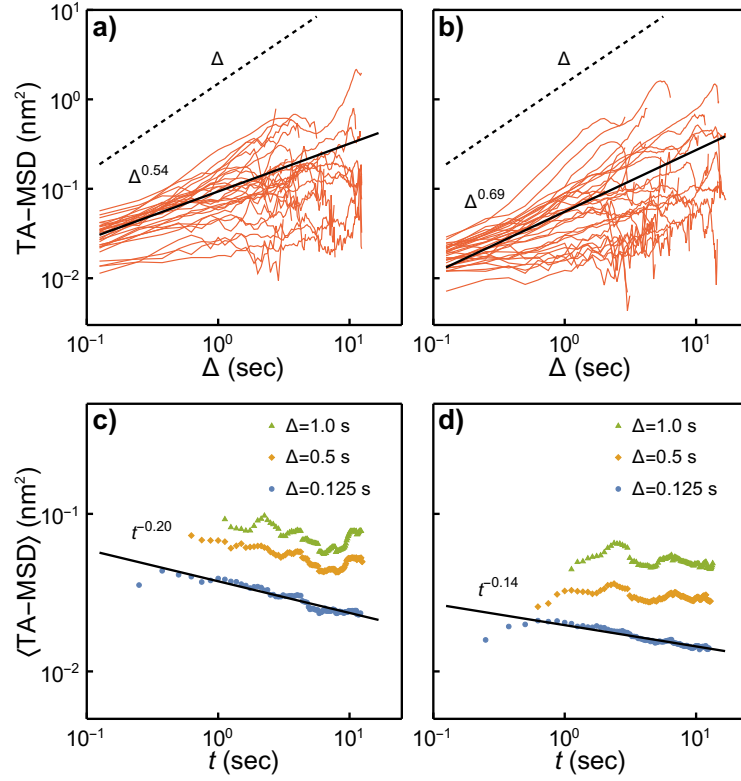
$$\langle \mathbf{r}^2(t) \rangle = 4D_\alpha t^\alpha \quad (5.4)$$

where  $\alpha$  is the anomalous diffusion exponent, and  $D_\alpha$  is the anomalous diffusion coefficient to distinguish it from  $D$ . When  $0 < \alpha < 1$  the phenomenon is known as *subdiffusion*, and when  $\alpha > 1$  it is known as *superdiffusion*. The rest of this chapter will focus on the subdiffusive regime, where the particle is diffusing slower than predicted by Brownian motion, since this corresponds to the measured exponents in Figures 5.6c and d. Superdiffusion on the other hand typically arises from a diverging second moment of  $P(\mathbf{r}, t)$ , leading to an increased probability of longer jumps. These are known as Lévy walks and flights, since this increased probability can be modelled using Lévy alpha-stable distributions [247]\*. An in-depth review of superdiffusion can be found in Ref. 248.

Returning to subdiffusion, numerous theories have been developed to explain the conditions in which anomalous behaviour arises. Common models include the continuous-time random walk (CTRW), where a particle jumps between traps with random dwell times [249]; random walks on fractals, with no characteristic length scales for particle displacements [247, 250]; and fractional Brownian motion (FBM), where the diffusing particle exhibits long-term correlations in displacements [251]. In complex, heterogeneous systems, several models, along with thermal noise, can combine to give rise to anomalous diffusion [252]. For an in-depth discussion of subdiffusion models, readers are referred to the recent review by Metzler and coworkers [19].

---

\*The normal distribution is in fact a special case of a Lévy alpha-stable distribution with  $\alpha = 2$ , and is the only one with a defined variance.



**Figure 5.8:** (a,b) TA-MSD of the trajectories extracted from sequence 1 and sequence 2 respectively. (c,d) Mean TA-MSD of the trajectories from each sequence as the observation time  $t$  increases. Best-fit lines are indicated in black with corresponding exponents, and the dashed lines show the behaviour expected from normal diffusion.

Alongside the non-linearity in Equation 5.4, another key consideration for anomalous diffusion is the *ergodic* behaviour of the system. The ergodic hypothesis states that, in the case of diffusion, the statistical properties of the whole ensemble of particles such as the mean and variance (i.e. MSD) can be deduced from a sufficiently long observation of a single particle. Ergodicity in such a system of diffusing particles can be assessed by considering the relationship between the ensemble MSD and the time-averaged MSD (TA-MSD), defined as:

$$\overline{\delta^2(\Delta)} = \frac{1}{t - \Delta} \int_0^{t-\Delta} \{\mathbf{r}(\tau + \Delta) - \mathbf{r}(\tau)\}^2 d\tau \quad (5.5)$$

in which  $\Delta$  defines the width of a sliding window along the trajectory of a particle. In an ergodic system,  $\langle \mathbf{r}^2(t) \rangle = \overline{\delta^2(\Delta)}$  as the measurement time  $t \rightarrow \infty$ . Non-ergodic behaviour cannot be measured solely from the ensemble averages provided by experimental techniques such as concentration profiles, and so single particle tracking as demonstrated here is therefore critical to understanding the physical models for atomic diffusion. Figures 5.8a and b plot the TA-MSD for each trajectory extracted from the two sequences, and two important points arise. First, the significant scatter

in amplitudes, which increases with  $\Delta$ , is a clear indication of ergodicity breaking. Secondly, the averaged exponents for each sequence differ from the corresponding MSD exponents (Figs. 5.6c and d), so the ergodic hypothesis evidently does not hold in this system.

A related quantity that be analyzed as a result of single particle tracking is the ensemble TA-MSD, which measures the general behaviour in the TA-MSD as the observation time  $t$  increases:

$$\langle \overline{\delta^2(\Delta)} \rangle = \frac{1}{N} \sum_{i=1}^N \overline{\delta_i^2(\Delta)} \quad (5.6)$$

The diffusive behaviour of many diffusion models does not depend on how long the system is observed for, so the ensemble TA-MSD is a useful method for determining which models might be applicable. Figures 5.8c and d show a power-law decrease in the ensemble TA-MSD for both sequences, which is indicative of a phenomenon known as *ageing*, whereby the observed diffusion behaviour depends on both the observation time and on the time that has passed since the system was created. In the present Cu/GO system, over time the Cu adatoms will tend to diffuse to stable pinning sites such as disordered, oxidized regions, and so the initial rapid diffusion over  $sp^2$  graphene will give way to slower diffusion in the disordered regions.

Using the ensemble TA-MSD, a further measure of ergodicity breaking can be defined as the parameter  $\xi$ , which represents the scatter seen in the TA-MSDs of each trajectory:

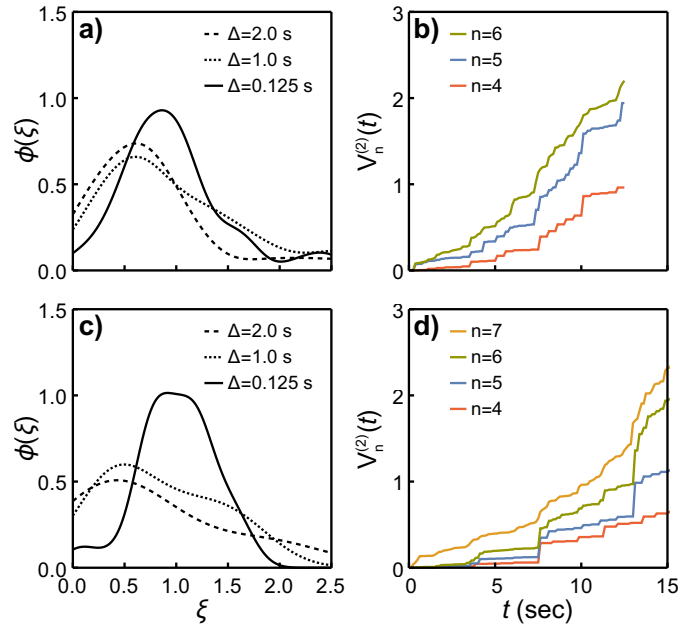
$$\xi = \frac{\overline{\delta^2(\Delta)}}{\langle \overline{\delta^2(\Delta)} \rangle} \quad (5.7)$$

The distribution  $\phi(\xi)$  shows a finite value at  $\xi = 0$  for both sequences (Figs. 5.9a and c), whereas an ergodic process would be expected to have a sharp peak at  $\xi = 1$ . The increasing scatter in the TA-MSDs is reflected in the increasing width of  $\phi(\xi)$  with  $\Delta$ . Another test of the anomalous diffusion behaviour uses the  $p$ -variation [253], which separates a particle trajectory into  $2^n$  segments, then the position differences between segments are summed and raised to the  $p^{\text{th}}$  power. The  $p$ -variation for  $p = 2$ ,  $V_n^{(2)}(t)$ , would be linear in  $t$  and increase in gradient with increasing  $n$  for an ergodic process, but the discrete jumps in Figures 5.9b and d again indicate that the experimental observations exhibit ergodicity-breaking.

## 5.5 A hybrid model for adatom diffusion on GO

Identifying the physical origins of the observed subdiffusion, and in particular the non-ergodicity, as well as why the behaviour in the two sequences should be subtly different, relies on understanding the models behind anomalous diffusion and how





**Figure 5.9:** Testing the origins of subdiffusivity on graphene oxide. **(a)** Distribution of scatter in TA-MSD for different intervals  $\Delta$  in sequence 1. **(b)** The  $p$ -variation ( $p = 2$ ) of one of the experimental trajectories from sequence 1. **(c)** Distribution of TA-MSD scatter for sequence 2. **(d)** The  $p$ -variation ( $p = 2$ ) of one of the experimental trajectories from sequence 2.

these relate to the experimental system. While a powerful tool for understanding structures at the atomic scale, DFT alone is unsuitable for modelling the dynamic behaviour of atoms in complex systems over long timescales. Accounting for the effects of the electron beam further complicates the calculations, and rapidly increases the computational resources required for a full first-principles simulation. Moving to classical MD methods extends the timescales that can be investigated, although realistically only as far as the nanosecond regime rather than the millisecond-scale observations in the experiments here. Furthermore, parameterizing empirical potentials with sufficient accuracy and complexity to model all the possible atomic interactions is not a straightforward task. Kinetic Monte Carlo (KMC) simulations can be linked with DFT calculations of transition pathways and energy barriers to simulate complex surface diffusion behaviours [254], but again including all of the possible interactions and sites in a heterogeneous substrate such as GO is limited by the need for extensive DFT simulations.

Stochastic simulations of diffusion offer a simpler approach to understanding the observed atom motion at a much-reduced computational cost. Any model must account for the variety of adatom binding sites, in terms of both relative stability and spatial distribution, as well as the changes in the substrate induced by the electron beam. The ergodicity breaking observed in the previous section rules out ergodic models such as FBM, but it is possible for multiple models to combine in the case of complex,

heterogeneous systems [21]. Knowing that GO exhibits both energetic disorder (in the Cu adatom binding sites) as well as spatial disorder (Fig. 5.4a), this section constructs a hybrid model for Cu adatom diffusion that is able to account for the two types of disorder present in GO.

### 5.5.1 Energetic disorder and the continuous-time random walk

Developed by Montroll and Weiss in 1965, the continuous-time random walk (CTRW) is one of the earliest models for describing anomalous diffusion [249]. It extends the stochasticity seen in the distribution of jump displacements to also incorporate random waiting times between jumps. This can be imagined as a particle moving on an energy surface containing a number of trapping sites of different depths (Figure 5.10a). The probability of the particle escaping a particular well is directly related to the depth of the well, and so deeper wells lead to longer waiting times before a particle is seen to move between sites. The mean time between particle jumps is now dependent on the waiting time PDF:

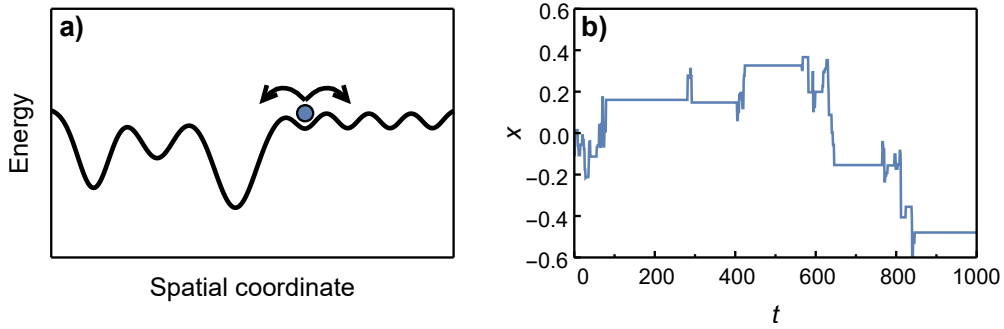
$$\langle \tau \rangle = \int_0^{\infty} \tau P(\tau) d\tau \quad (5.8)$$

The typical model used for the PDF is a power-law:

$$P(\tau) = \begin{cases} 0 & \tau < \tau_0 \\ \frac{\gamma \tau_0^\gamma}{\tau^{1+\gamma}} & \tau \geq \tau_0 \end{cases} \quad (5.9)$$

where  $\gamma$  is the power-law exponent. Here  $\tau_0$  is a parameter controlling the minimum possible waiting time relative to the laboratory time  $t$ . For example, if observations are made every  $t$  timesteps, and  $\tau_0 < t$ , this means that more than one jump can take place between observations. This is a critically important consideration in the context of time-resolved STEM imaging, which takes place over a much longer timescale ( $10^{-1}$  s) than the timescale of atomic motion between two neighbouring sites (ca.  $10^{-9}$  s) and so only captures the “before” and “after” states of the system.

Figure 5.10b shows a one-dimensional CTRW, with normally-distributed displacements. The effect of the power-law waiting times is clear, with periods of mobility separated by long periods of particle trapping. The periods of immobility lead to the system exhibiting ergodicity breaking, since it is now possible for particles to remain pinned in one site for the entire experiment. Analysis of the TA-MSDs of these immobile particles clearly will not reflect the ensemble MSD behaviour, and so the CTRW is a non-ergodic process. Immobile Cu atoms are observed experimentally on the GO substrate; Figure 5.5 uses them to align the image sequences and remove the effects of sample drift. The reason for the trapping is likely a result of a particularly strong binding site (e.g. due to oxygen or a graphene vacancy).



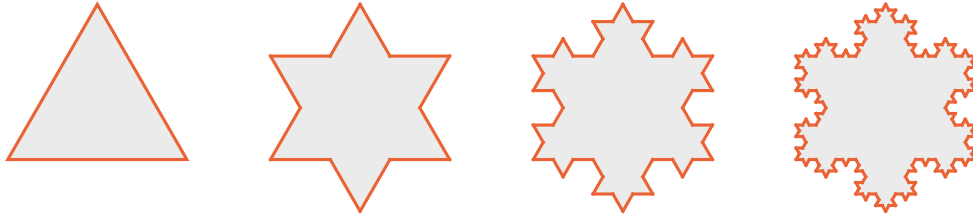
**Figure 5.10:** (a) Schematic diagram of a continuous-time random walk. (b) One-dimensional continuous-time random walk. Waiting times are drawn from a power-law distribution with  $\gamma = 0.8$  and  $\tau_0 = 1$ . Displacements are drawn from a normal distribution.

Previous theoretical analysis of the kinetics of adatom motion under an electron beam used elastic scattering cross-sections to determine rates of motion in comparison to thermal displacements [101]. As well as the electron beam geometry and specimen tilt, the cross-section also depends on the energy barrier for diffusion, which in the case of a heterogeneous substrate such as GO can be highly variable due to the arrangement of functional groups. This variability motivates the use of a CTRW model to account for the energetic disorder.

Another particular property of a CTRW is that these periods of immobility will increase as the system develops over time. Particles exploring the landscape will discover ever-deeper traps, leading to the phenomenon known as ageing. This can be characterized by the behavior of the ensemble TA-MSD of a CTRW in the limit  $\Delta \ll t$ :

$$\langle \overline{\delta^2(\Delta)} \rangle \propto \frac{\Delta}{t^{1-\gamma}} \quad (5.10)$$

Figures 5.8c and d show the expected decay of  $\langle \overline{\delta^2(\Delta)} \rangle$  with  $t$ , indicating that the system does exhibit ageing. However, the relationship between these exponents and the MSD exponents (Fig. 5.6c,d) does not show the expected behaviour from a pure CTRW. Furthermore, the TA-MSD of a CTRW should be linear in  $\Delta$ , but this is not seen experimentally. These discrepancies point to an important physical characteristic of the system that is not incorporated into the CTRW model. Although it incorporates the energetic disorder of GO uncovered by DFT calculations, the pure CTRW model was originally developed for walks on a regular lattice, or with the displacements drawn from a normal distribution. It is clear from the experimental observations that this is not an appropriate model for GO, and thus incorporating spatial disorder may explain the discrepancies in the exponents.



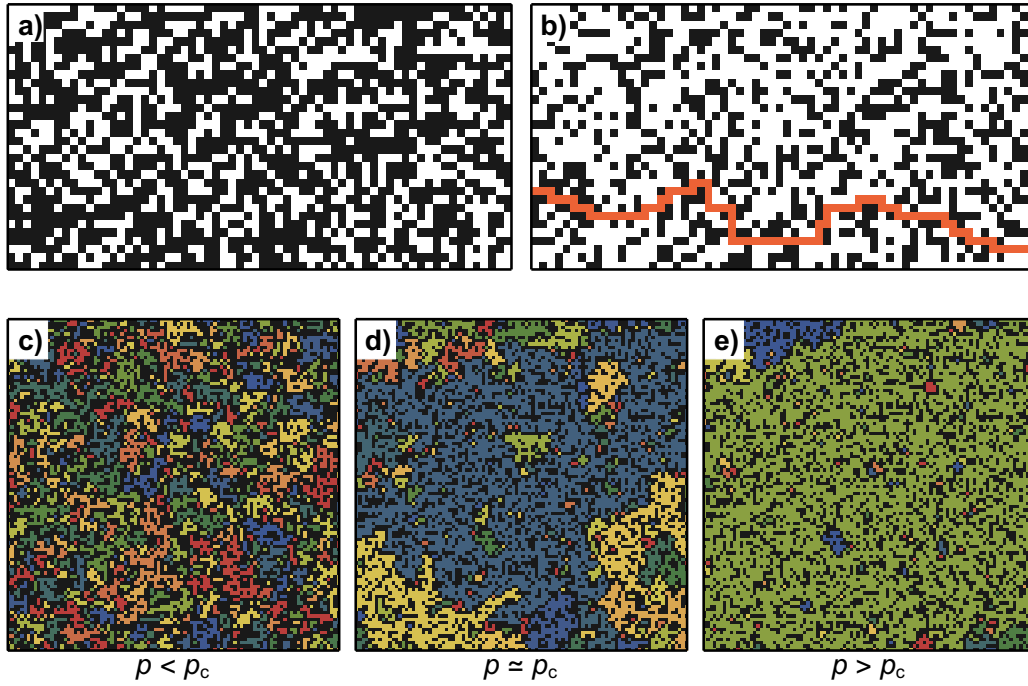
**Figure 5.11:** Starting from an equilateral triangle, the first four iterations of the Koch snowflake are shown. While the area converges to  $1.6\times$  the area of the starting triangle, the perimeter diverges to infinity. The Hausdorff dimension  $d_f$  is  $\log 4/\log 3 \approx 1.26$ .

### 5.5.2 Spatial disorder and percolation theory

Modelling spatially-disordered systems is often performed using the framework of percolation theory [250], an elegant approach that brings together the mathematics of fractals with experimental observations in an enormous range of physical systems. Fractals, first introduced by Mandelbrot [247], are defined mathematically as objects with a non-integer Hausdorff dimension,  $d_f$ . Fractals also exhibit a property known as *self-similarity*: at any magnification, the object looks exactly the same. *Deterministic* fractals are generated by recursively applying a set of operations, and include classic fractals such as the Sierpinski gasket and the Koch snowflake, with the latter being an example of how a finite area can be bounded by an infinitely long line (Fig. 5.11). While  $d_f$  can be defined exactly for deterministic fractals, many other systems can exhibit the same self-similarity property. These are known as *statistical* fractals, and their mass  $M$  scales with their linear size  $L$  as  $M \propto L^{d_f}$ . Mandelbrot identified that many systems in nature can be described as statistical fractals, including, famously, the coastline of Britain ( $d_f \approx 1.25$ ) [255, 256].

Percolation theory is a common framework for systems exhibiting statistical fractal character. The present discussion focuses only to the site percolation model, as opposed to bond percolation, since the bond model can be reformulated as a site model on a different lattice. Figure 5.12 depicts a square site lattice, where a fraction of sites  $p$  are occupied (white) and  $1-p$  are unoccupied (black). Connected percolation clusters are formed by occupied nearest neighbours, and as  $p$  increases so too does the average size of the clusters. Above the critical fraction,  $p_c$ , there exists a connected cluster extending from one side of the lattice to the other (Fig. 5.12b), allowing continuous transport (or “percolation”) across the lattice. This connected cluster is also known as an “infinite” cluster, and it has been shown that for a 2D site lattice  $d_f = 91/48 \approx 1.896$  [250].

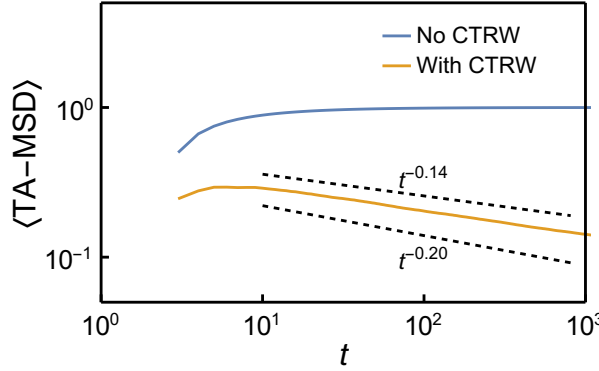
Examples of experimentally-observed percolation clusters include discontinuous thin films of Pb and Au [257, 258]. More recently, measurements of the electrical conductivity of GO at various levels of oxidation suggest that the spatial heterogeneity in



**Figure 5.12:** An example of a square site percolation lattice ( $p_c = 0.5927\dots$ ). **(a)** When  $p < p_c$ , there is no continuous path from the left-hand side of the lattice to the right. White squares are occupied sites, black squares are unoccupied. **(b)** When  $p \geq p_c$ , it is now possible to travel from left to right in a continuous path. The shortest such path is highlighted in orange. **(c)** When  $p < p_c$ , the lattice is characterized by many small unconnected clusters of accessible sites. Each colour denotes a connected cluster of occupied sites, and black represents unoccupied sites. **(d)** When  $p \simeq p_c$ , a few larger connected clusters begin to dominate the lattice, although some small clusters remain. **(e)** When  $p > p_c$ , a single cluster now covers most of the lattice.

GO can be described using percolation theory [259]. The  $sp^3$  character introduced by functional groups disrupts the conductivity of GO relative to pristine graphene. As the oxygen ratio is reduced, the fraction of GO with  $sp^2$  character increases, until it forms connected clusters enabling excellent conductivity across the GO sheet. This is demonstrated in Figure 5.12c-e, where the increase in the occupancy (i.e. the ratio of  $sp^2$  to  $sp^3$  regions) leads to the formation of connected clusters.

Turning now to the case of a random walk on a percolation cluster, the spatial disorder and lack of a characteristic length scale results in anomalous diffusion behaviour [250]. An important distinction is between behaviour below and above the percolation threshold at which the infinite connected path of accessible sites is formed, as diffusive behaviour transitions from non-ergodic on separate, isolated clusters to ergodic on the infinite cluster. However, random walks on a percolation cluster do not exhibit the ageing of a CTRW that is seen experimentally, suggesting that a hybrid model combining both spatial and temporal disorder is more appropriate for the metal-GO system.



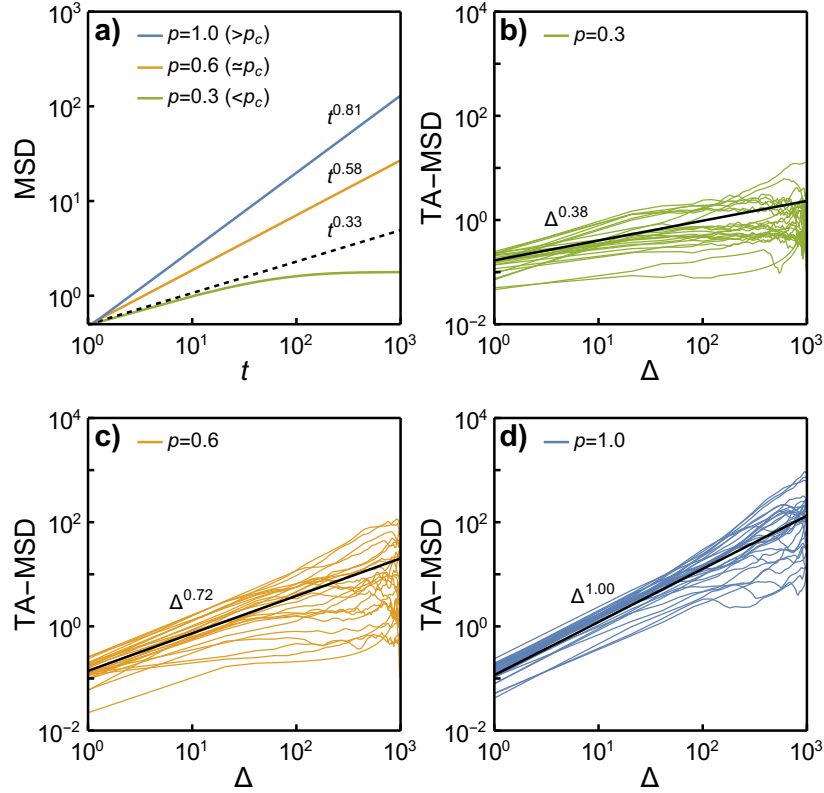
**Figure 5.13:** Comparison of the ageing behaviour ( $\Delta = 1$ ) of a random walk on a honeycomb site lattice with occupation  $p = 0.6$ , with and without a CTRW contribution ( $\gamma = 1$  and  $\tau_0 = 1$ ). The dashed lines are the measured slopes for sequences 1 and 2.

### 5.5.3 The hybrid model

The hybrid model for adatom diffusion on GO was developed from previous works combining diffusion models [21, 260], based on the percolation cluster representation of GO described above. First, percolation on a  $256 \times 256$  honeycomb site lattice with periodic boundary conditions was simulated. Random walks over the accessible sites of the lattice were then generated. The durations of each step were drawn from a power-law distribution (Eq. 5.9). Only those steps occurring within a given observation time  $t$  were kept to form the hybrid model. The possibility of missed jumps in the experimental system, where an atom moves through several sites in between frames, can be incorporated by choosing  $\tau_0 < 1$ .

It was suggested previously that random walks on isolated clusters in a percolation lattice ( $p < p_c$ ) give rise to non-ergodic behaviour without the need for a CTRW contribution. Indeed, the low probability of the existence of an infinite path of accessible sites on GO (cf. Fig. 5.4) suggests a value of  $p$  below  $p_c$ , which for a honeycomb lattice is 0.697. However, as shown in Figure 5.13, a random walk on these clusters also requires the incorporation of a CTRW to model the ageing behaviour seen in the experimental system.

Figure 5.14 considers further the effect of the occupation fraction  $p$  on the diffusion behaviour, since this relates to the number of available sites for pinning a Cu atom to GO. For low availability ( $p \ll p_c$ ), the atoms are confined to small clusters and are unable to fully explore the lattice, so the MSD exhibits a plateau at long times (Fig. 5.14a), similar to the behaviour seen in the MSD for sequence 1. For high availabilities the fractal contribution to the random walk is negligible, and so a pure CTRW model dominates with the TA-MSD being linear with respect to  $\Delta$ .

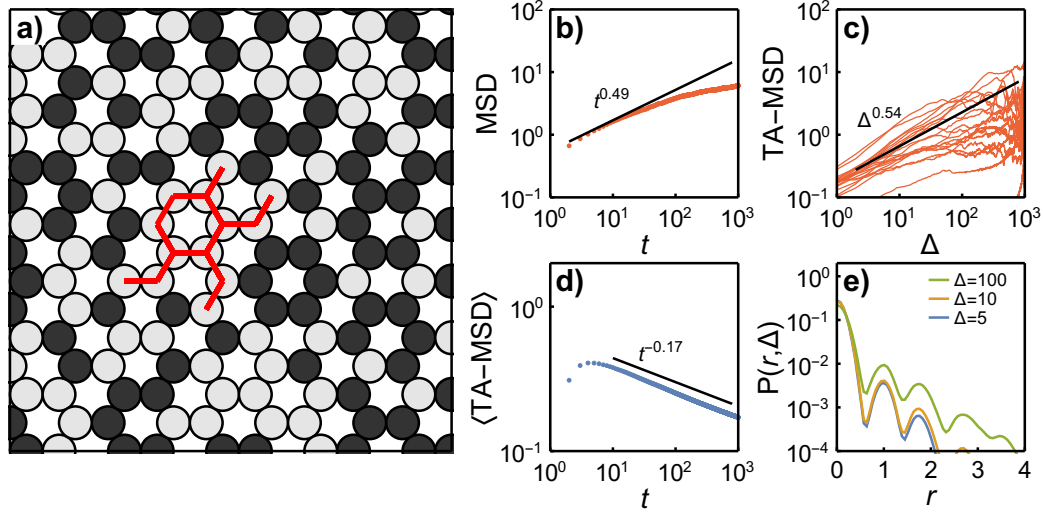


**Figure 5.14:** The effect of the occupation fraction  $p$  on the hybrid model. **(a)** MSD of  $10^4$  CTRWs with  $\gamma = 0.8$  and  $\tau_0 = 1$  on a honeycomb site lattice ( $p_c = 0.697$ ) with varying occupancies. **(b-d)** Subset of TA-MSDs for the same walks. The solid lines are fits to the simulated TA-MSDs.

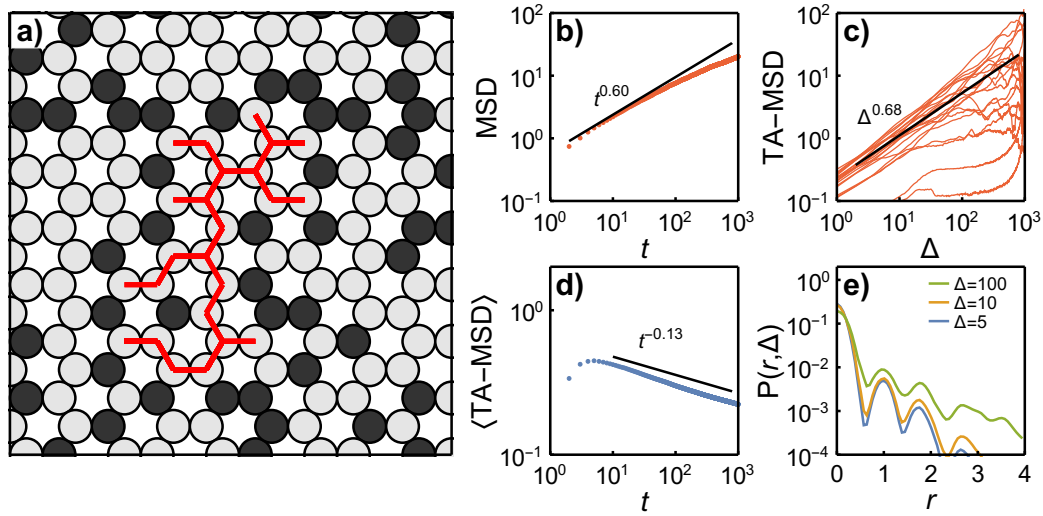
Combining the exponents from the experiment with the results from the simulations, the hybrid model parameters ( $p, \gamma, \tau_0$ ) were fitted using a least-squares minimization in the absence of measurement noise. For the model of sequence 1, the fraction of accessible sites was  $p = 0.45$ , with power-law exponent  $\gamma = 0.88$ , and  $\tau_0 = 0.70$ . For sequence 2, the fraction of accessible sites increased to  $p = 0.60$ , the power-law exponent increased to  $\gamma = 0.98$ , and  $\tau_0 = 0.70$ .

Figures 5.15 and 5.16 show the behaviours of the two fitted hybrid models. The schematic diagrams show a simulated random walk on a honeycomb site lattice without measurement noise, while plots of the MSD, TA-MSD, ensemble TA-MSD and jump displacement distributions show how the models accurately reflect the experimental observations. For these plots, a small amount of Gaussian noise was added to the trajectories to reflect the uncertainties in measuring the positions of atoms experimentally; more information about noise is given in Section 5.5.4.

When a particle arrives at a lattice site, the waiting time before the next jump is drawn from a power-law distribution as in the CTRW model. Particles revisiting a site will have a new waiting time, thus reflecting any beam-induced changes in the substrate that



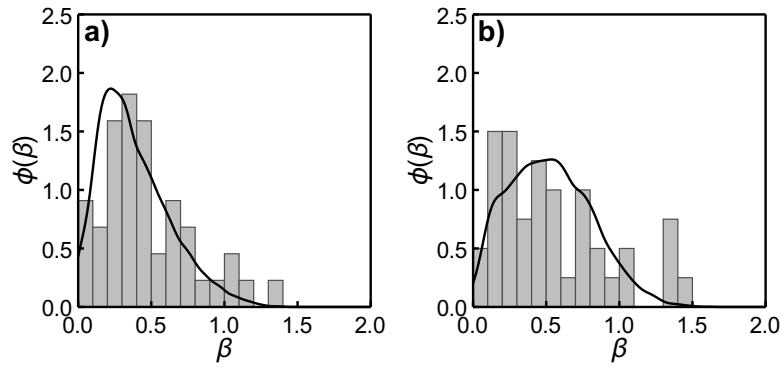
**Figure 5.15:** A model for anomalous diffusion of adatoms on graphene oxide in sequence 1. (a) Honeycomb site lattice with probability of occupancy,  $p = 0.45$ ; grey circles represent accessible sites, while black circles represent inaccessible sites. The red line shows a CTRW on a cluster of accessible sites. (b) Ensemble MSD, (c) TA-MSD, (d) mean TA-MSD for  $\Delta = 1$ , and (e) jump length distribution of simulated CTRWs on the lattice with  $\gamma = 0.88$ ,  $\tau_0 = 0.70$ , and a small amount of Gaussian noise added.



**Figure 5.16:** A model for anomalous diffusion of adatoms on graphene oxide in sequence 2. (a) Honeycomb site lattice with probability of occupancy,  $p = 0.60$ ; grey circles represent accessible sites, while black circles represent inaccessible sites. The red line shows a CTRW on a cluster of accessible sites. (b) Ensemble MSD, (c) TA-MSD, (d) mean TA-MSD for  $\Delta = 1$ , and (e) jump length distribution of simulated CTRWs on the lattice with  $\gamma = 0.98$ ,  $\tau_0 = 0.70$ , and a small amount of Gaussian noise added.

may affect the copper binding energy. This also accounts for the stochastic nature of the electron–nucleus collision events. The honeycomb lattice incorporates the underlying



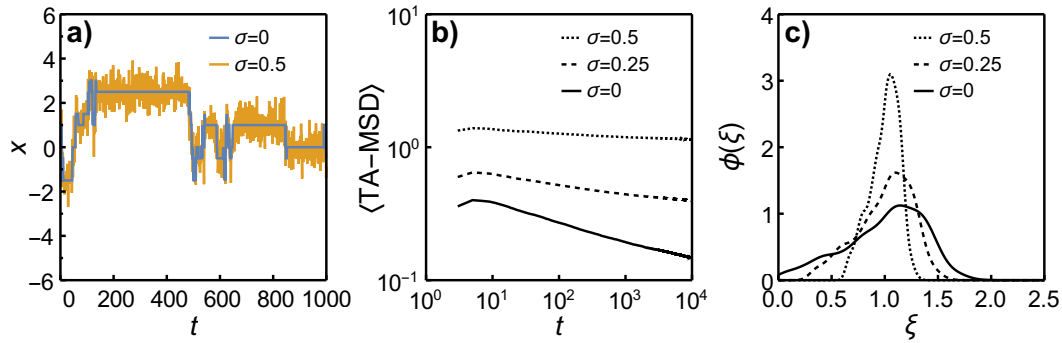


**Figure 5.17:** Distribution of TA-MSD exponents,  $\beta$ . **(a)** Sequence 1 (grey bars) and the hybrid model (solid line). **(b)** Sequence 2 (grey bars) and the hybrid model (solid line).

hexagonal symmetry of the GO, and the possibility of five- and seven-membered rings in defective GO regions is accounted for by the structural disorder introduced by the site percolation aspect of the model. Using this hybrid model we match remarkably well the observed subdiffusive behaviour seen in the MSD (Fig. 5.15b), including a long-time plateau due to confinement (cf. Fig. 5.6c). Ergodicity breaking is identified by the sub-linear behaviour and scatter in the TA-MSD (Fig. 5.15c), the ageing behaviour (Fig. 5.15d) and the characteristic jump probabilities (Fig. 5.15e).

Further confirmation of the fit between the hybrid model developed here and the experimental observations is shown in the scatter of the TA-MSDs (Fig. 5.17). The distribution of TA-MSD exponents,  $\phi(\beta)$ , shows good agreement between the experimental results and the simulations for both sequences, with a finite value at  $\beta = 0$  indicating completely immobile atoms.

The hybrid model provides a clearer understanding of the physical processes involved in the anomalous diffusion, and allows us to return, to the Cu–GO system and explain the non-ergodic subdiffusive adatom motion in more detail. The electron beam provides sufficient energy for Cu atoms to explore a distribution of binding sites with a range of energies. Adatom movement on  $sp^2$  graphene will be rapid as the binding energies are small. The important binding sites, however, corresponding to the locations of oxygen atoms bound to the underlying graphene, are spatially disordered and can be modelled as a percolation network, and the ageing corresponds to adatoms encountering ever-deeper trapping sites. Although the characteristic behaviour of the adatom motion in sequences 1 and 2 is similar, the subtle differences in ageing and the degree of ergodicity-breaking can be explained by a beam-induced change in the GO support. This could be due to an increase in oxide coverage or hydrocarbon contamination. The change leads to a reduction in the area of  $sp^2$  graphene, and thus a reduction in the number of weak binding sites available, and a corresponding increase in the fraction of stronger binding sites to which the adatom may be pinned. Other changes such as the



**Figure 5.18:** Assessing the effect of measurement noise on the hybrid model. **(a)** The  $x$  component of a CTRW on a honeycomb site lattice without ( $\sigma = 0$ ) and after ( $\sigma = 0.5$ ) the addition of Gaussian noise. The distance between adjacent lattice sites is 1. **(b)** Behaviour of the mean TA-MSD as the measurement noise level increases. **(c)** Distribution of the scatter in TA-MSD ( $\Delta = 5$ ) for increasing measurement noise.

formation of graphene vacancies and the presence of impurity atoms will only increase the variety of available binding sites, further motivating the use of a stochastic CTRW model to account for the adatom binding.

#### 5.5.4 Accounting for measurement noise

Errors in the position and tracking of atoms in the sequences can lead to the observation of anomalous diffusion behaviour even in the case of normal Brownian motion, so as a final check it is important to rule out the effects of measurement noise. The Gaussian fits in Figure 5.7a and b provide an estimate of the measurement noise in the system, since the small particle displacements most likely correspond to fluctuations in the recorded position of an atom in the same site. These are either a result of thermal displacements or errors in the atom-fitting procedure. For the interval  $\Delta = 0.125$  s, the full-width half-maximum of the Gaussian fit was 0.09 nm for sequence 1 and 0.07 nm for sequence 2, compared to the graphene C–C bond length of 0.142 nm. Using this information, random walks drawn from the hybrid model were combined with additive Gaussian noise. Only when the additive noise becomes significant relative to the size of the percolation lattice are the ageing and ergodicity breaking identified in the experimental system masked by the noise (Fig. 5.18). This confirms that the observations do not arise from measurement noise but are in fact an inherent property of the system.

## 5.6 Summary

Using fast atomic resolution STEM imaging in combination with the denoising and particle tracking approaches discussed previously, this chapter has investigated the diffusive motion of Cu adatoms on graphene oxide. The system was found to exhibit anomalous subdiffusive behaviour, and this was explained using a hybrid CTRW model combined with a site percolation network. The development of a simple, stochastic model provides a flexible method for simulating the interactions of atoms with imperfect surfaces, where the complexity makes DFT simulations computationally demanding. Such a model paves the way towards a more comprehensive understanding, and prediction, of atomic-scale surface diffusion in materials systems possessing spatial and energetic heterogeneity. Furthermore, it provides a route for improving our understanding of the activity, stability and durability of single atom catalysts.

One of the questions that arises from the observations and analyses of this chapter is why do the Cu adatoms remain fairly well-separated, and why is no significant agglomeration into small atomic clusters seen? Indeed, one of the assumptions in the hybrid model developed here is that the Cu atoms are largely independent of each other, but is that truly the case? The atoms were initially deposited as Cu<sub>13</sub> clusters, but these appear to have broken up prior to observation in the STEM. Accounting for this behaviour is an important next step to characterizing single atom and atomic cluster catalysts, since the effect of adding or removing just one atom on the catalytic activity can be significant. The next chapter will present a more in-depth DFT study of the behaviour of Cu atoms on graphene and GO substrates, with extensive use of the AIRSS technique to explore and understand the complexity of the system.



## Chapter 6

# Structures of small copper clusters on graphene oxide

### 6.1 Introduction

The original decision to study the copper samples in the previous chapter was driven by the fact that small copper clusters are becoming increasingly important for catalysis [37]. Although the fabrication method involved the soft landing of size-selected  $\text{Cu}_{13}$  clusters onto the GO substrate, very little evidence of  $\text{Cu}_{13}$  was found upon investigation using STEM. Instead the Cu was largely dispersed as single atoms over the surface, as seen in Chapter 5. Preferential binding to oxygen-rich disordered regions was observed, and this raises the question of whether the oxygen functional groups, originally intended to pin the Cu clusters, have instead contributed to their fragmentation. Previous work on other substrates suggests that the soft landing fabrication method alone does not typically lead to cluster fragmentation [261], so investigating other possible causes is key to a full understanding of the cluster–substrate interaction and the system as a whole. Crucially, although small clusters are often more active catalysts than larger nanoparticles, the decreased stability often results in an early decrease in catalytic activity as the particles fragment or agglomerate [36].

This chapter reanalyzes the samples from Chapter 5, although the focus is now on the relationships between copper atoms as well as the substrate\*. Evidence of small clusters, most notably a  $\text{Cu}_3$  motif, reveals anomalously large Cu–Cu interatomic distances on the substrate, indicating that the GO plays a critical role in stabilizing the adatoms.

---

\*The work presented in this chapter was carried out in collaboration with R.K. Leary, G. Schusteritsch, C.J. Pickard, E.C. Tyo, S. Vajda, J.M. Thomas, Q.M. Ramasse, P.D. Bristowe and P.A. Midgley. R.K.L. and Q.M.R. performed the ADF-STEM experiments. E.C.T. and S.V. prepared the copper samples. G.S. and C.J.P. assisted with the DFT and AIRSS calculations. The use of spatial point processes to analyze adatom distributions (Section 6.3) has also been published in Ref. 262.

Quantitative analysis of the adatom distributions is used to further understand these interatomic distances. In the second half of the chapter, a detailed computational investigation of the structure and stability of Cu clusters in vacuum, in the presence of oxygen, and on the surface of GO is conducted. This study develops the AIRSS simulations first used in Chapter 5 to identify the role played by oxygen in binding Cu atoms and clusters to the substrate.

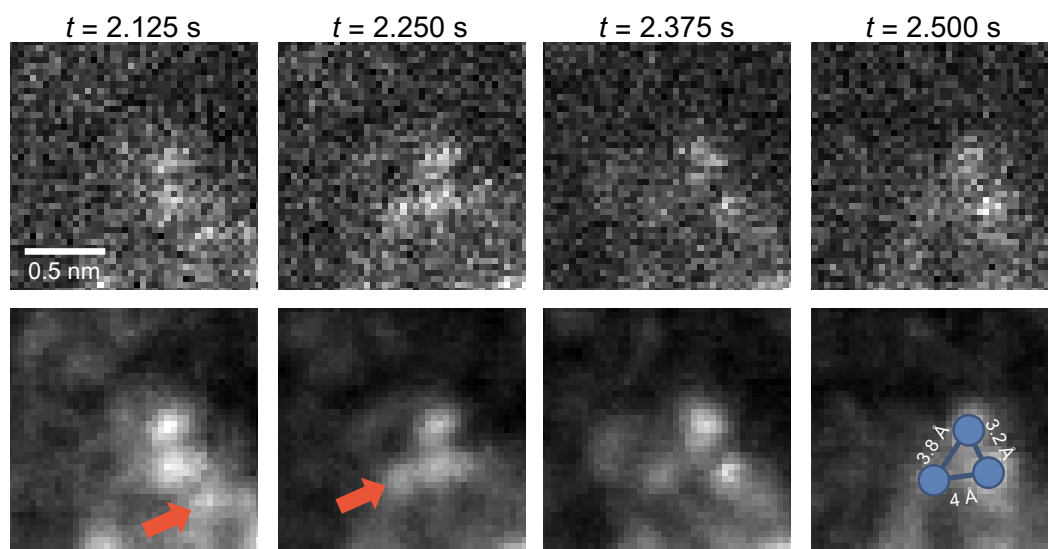
## 6.2 ADF-STEM results

As in Chapter 5, ADF-STEM imaging was carried out using two aberration-corrected microscopes. Static and time-resolved imaging of Cu clusters was performed on a FEI Titan<sup>3</sup> operating at 80 kV, and a Nion UltraSTEM 100 operating at 60 kV was used to acquire atomic-resolution images of the GO substrate.

### 6.2.1 Time-resolved observations

Figure 6.1 presents four consecutive frames of an image sequence acquired at 8 fps. The imaging conditions are largely identical to the sequences studied in Chapter 5, and the motion of Cu adatoms is again driven by the electron beam. The smaller pixel size in this sequence means that the dose rate is approximately twice that of the sequences in Chapter 5, i.e.  $10^7 \text{ e}^- \text{ nm}^{-2} \text{ s}^{-1}$ . Figure 6.1 shows an arrangement of adatoms before and after denoising with PGURE-SVT, starting 2 s after the area was first irradiated by the electron beam.

The first frame shows two pinned adatoms and a third, more mobile atom highlighted by the arrow. In the next frame, this third atom has jumped, eventually forming an asymmetric triangle motif with the pinned atoms by the third frame. This “trimer” motif remains stable under the beam for a further 1.5 s before the third atom disappears from the field-of-view (presumably sputtered off the surface), leaving behind the original pair of stable atoms. The third atom is evidently less strongly-bound to the substrate than the others, supporting the idea of a distribution of trapping sites for adatoms as outlined in Chapter 5. The projected interatomic spacings of the trimer motif are measured in the fourth frame of Figure 6.1, and it is immediately clear that the spacings are too large to be direct Cu–Cu bonds; the nearest-neighbour distance in fcc Cu is 2.55 Å by comparison. Clearly the substrate has a significant impact on the clustering behaviour of Cu, but it is difficult to fully characterize and understand this effect without better images of the underlying substrate. The role of oxygen pinning the atoms to the GO, first observed in the previous chapter, is of particular importance.



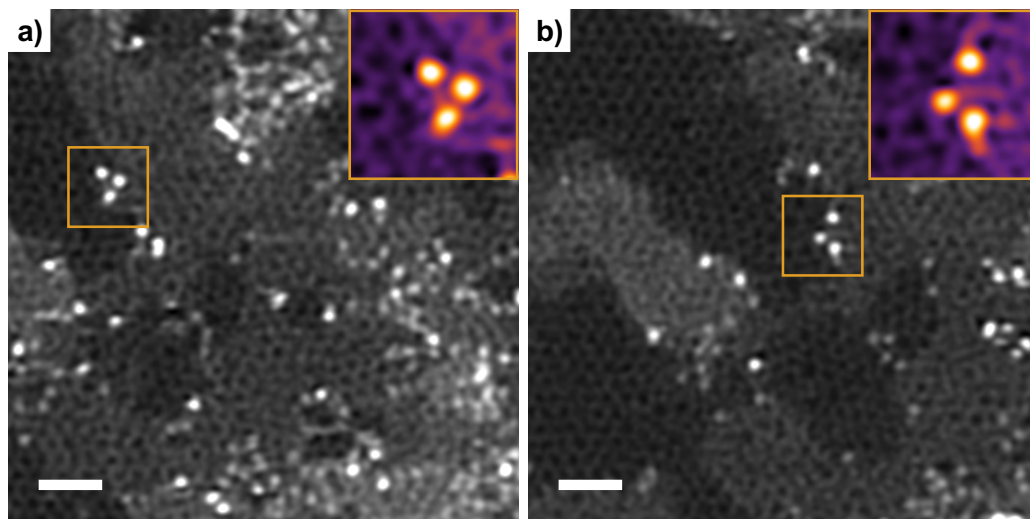
**Figure 6.1:** Consecutive noisy (top row) and denoised (bottom row) ADF-STEM frames of Cu adatoms on graphene oxide. At  $t = 2.125$  s, two adatoms are pinned in the centre of the frame, and a third atom is marked by the arrow. At  $t = 2.250$  s, the third atom moves as highlighted by the arrow. Finally at  $t = 2.375$  s, a stable trimer configuration forms, with interatomic distances measured at  $t = 2.500$  s. Figure adapted from Ref. 117. See also Movie S13.

### 6.2.2 Resolving the GO substrate

Turning now to the 60 kV ADF imaging carried out on a Nion UltraSTEM, Figure 6.2 presents two examples of Cu atoms on GO following baking at 120°C to drive off carbonaceous debris. These images are similar to Figure 5.1, but are important because they each contain an asymmetric trimer motif similar to Figure 6.1 (see insets). Here the projected interatomic spacings are measured to be slightly smaller than in the time-resolved example, at  $2.75 \pm 0.14$  Å,  $3.00 \pm 0.20$  Å and  $3.80 \pm 0.26$  Å. Despite the discrepancy with Figure 6.1, the spacings are still incompatible with a direct Cu–Cu bond. Furthermore, the improved resolution means that the relationship between the Cu cluster and the GO can be analyzed in greater detail. Both examples are seen to sit at the edges of disordered,  $sp^3$  regions on GO, as per the previous observations in Chapter 5. Filtering out the effects of probe tails reveals increased intensity *between* the Cu atoms that could conceivably be attributed to oxygen atoms, although these are too mobile under the beam to be precisely located and identified.

### 6.2.3 Estimating the fractal dimension of GO

While the results presented above focus on either the dynamics of adatoms or resolving the GO with atomic resolution, complementary information about the structure of GO can be ascertained by imaging the sample at larger length scales. Two ADF-STEM images (Fig. 6.3) were acquired on an FEI Titan<sup>3</sup> at an accelerating voltage of 80 kV

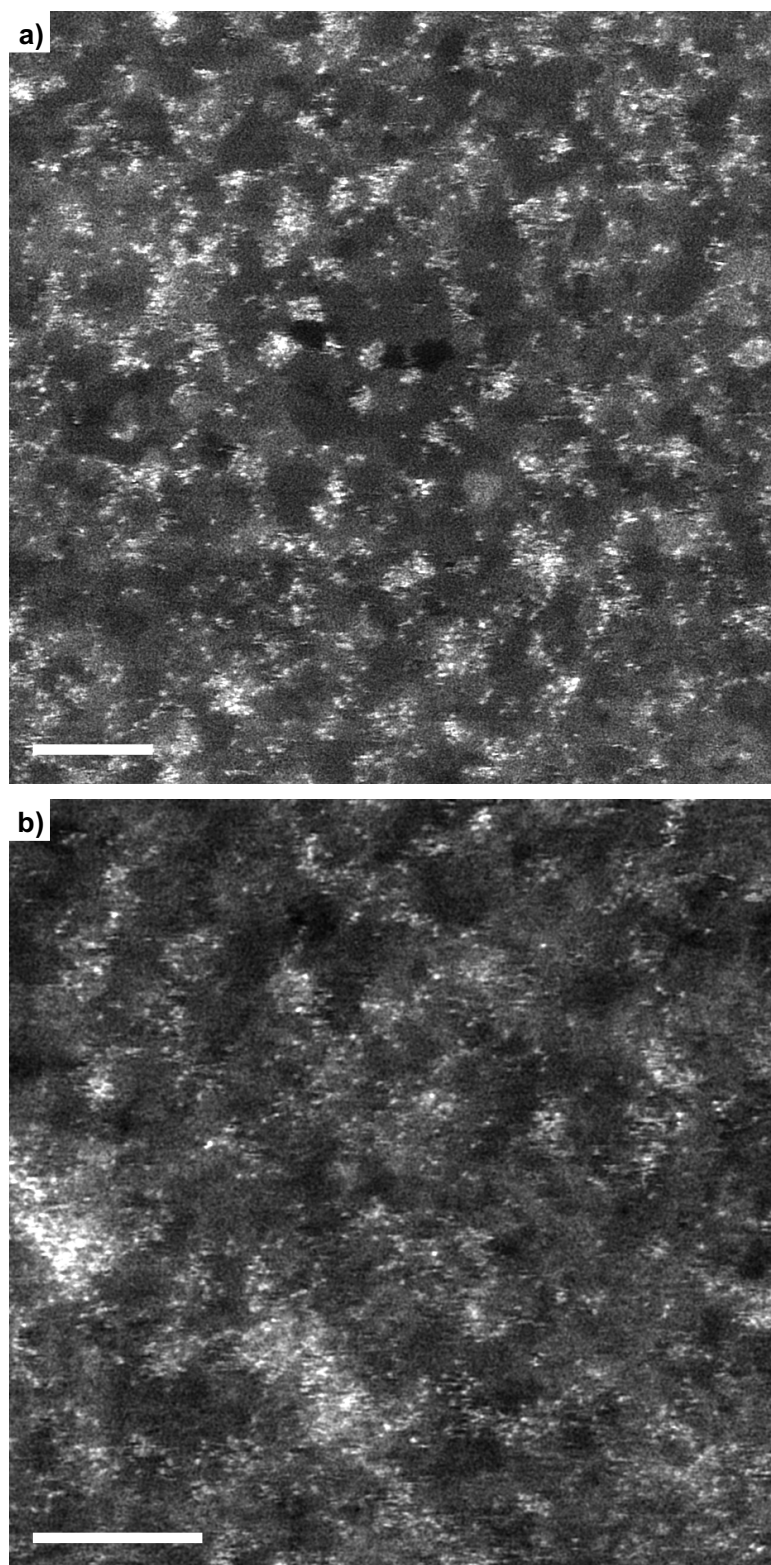


**Figure 6.2:** Two ADF-STEM images of Cu adatoms on GO, acquired at an operating voltage of 60 kV. The images have been filtered with a double-Gaussian filter to remove the probe tails [50]. The insets highlight two asymmetric trimer motifs of Cu adatoms. Scale bars: 1 nm.

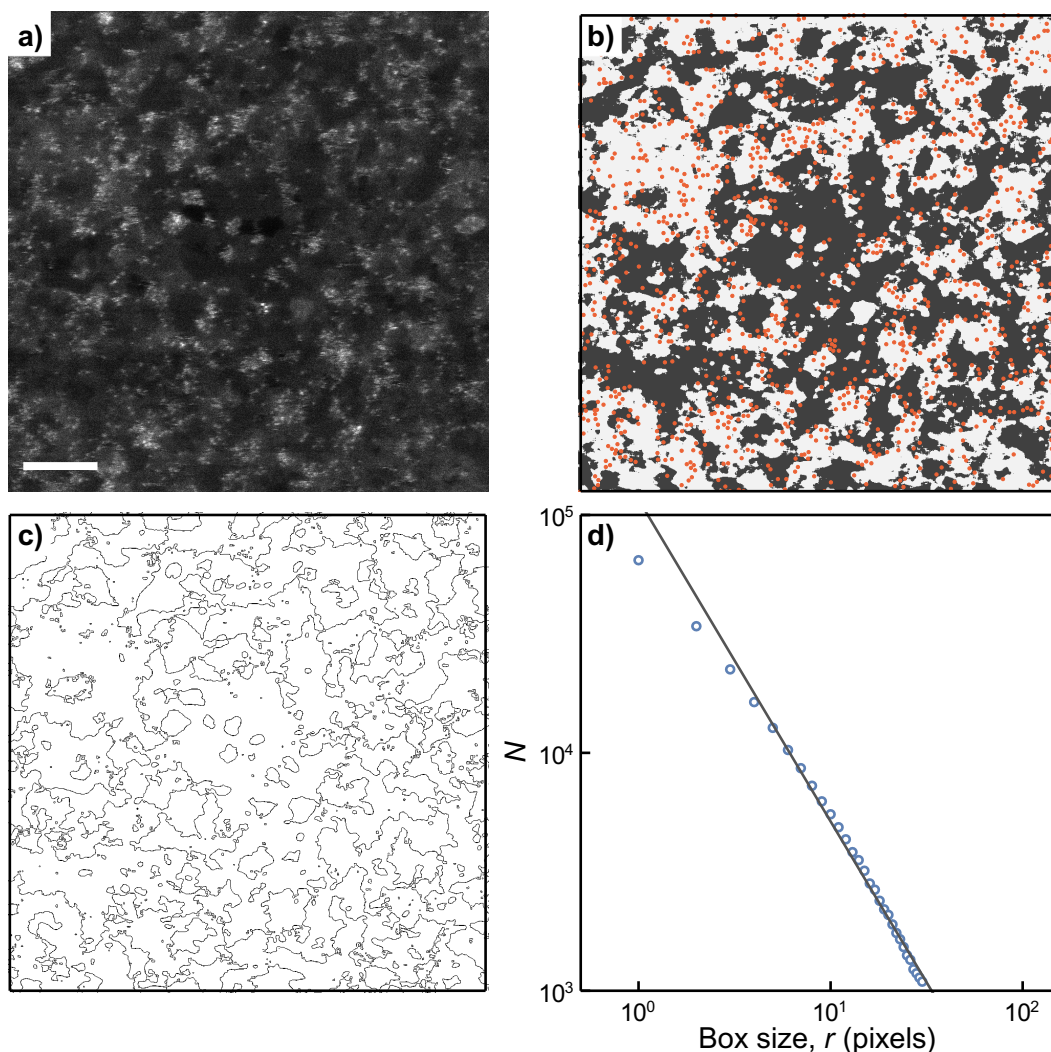
and with a probe current of ca. 60 pA. The pixel dwell-time and dose was similar to the time-resolved experiments, with the total dose estimated as  $10^6 e^- \text{ nm}^{-2}$ . The ADF detector inner angle was 60 mrad, rather than 36 mrad used in the time-resolved imaging, which significantly enhances the visibility of the Cu atoms relative to the GO substrate.

After applying an edge-preserving bilateral filter to smooth the images [263], a straightforward binary thresholding operation successfully separates the disordered and ordered regions of the GO substrate. The resulting binary image (Fig. 6.4b) looks remarkably similar to the percolation clusters of the previous chapter (cf. Fig. 5.12), giving further weight to the choice of a percolation model for explaining the anomalous diffusion exhibited by Cu atoms under the electron beam. The boundary between ordered and disordered regions, where adatoms are typically found, is obtained through the application of an edge detection filter (Fig. 6.4c). By using the box-counting method [264], which tallies the number of boxes of size  $r$  containing black pixels (representing edges), the fractal dimension  $d_f$  of the GO can be estimated (Fig. 6.4d). For both images in Figure 6.3,  $d_f$  is found to be ca. 1.59, compared to a theoretical value of 1.896 for the infinite percolation cluster at criticality. These values are in agreement with the estimate that  $p < p_c$  made in Chapter 5 using an analysis of the anomalous diffusion behaviour [265].





**Figure 6.3:** Two ADF-STEM images of Cu adatoms on GO, acquired at an operating voltage of 80 kV. The images have been filtered with a double-Gaussian filter to remove the probe tails [50]. Scale bars: 5 nm.



**Figure 6.4:** Determining the fractal dimension of GO. **(a)** An ADF-STEM image of Cu adatoms on GO. Scale bar: 5 nm. **(b)** Bilateral smoothing and binarization separates the image into ordered (grey) and disordered regions (white). The positions of Cu atoms extracted from **a** are overlaid onto the segmented image in orange, showing the preferential segregation of adatoms to disordered regions. **(c)** Applying an edge-detection filter highlights the boundaries between ordered and disordered regions. **(d)** A log-log plot of the number,  $N$ , of boxes of size  $r$  containing black pixels in **c** is used to determine the fractal dimension.

### 6.3 Distribution of copper adatoms on GO

The large field-of-view images in Figure 6.3 also provide a useful dataset for characterizing the distribution of Cu adatoms over the surface of GO. A Laplacian-of-Gaussian (LoG) filter, as implemented in the scikit-image toolbox [266], was used to extract the positions of Cu atoms in each image. A LoG filter was chosen here over the alternative methods presented in Chapter 4 for a number of reasons. Alongside computational efficiency, the atoms in these images have a higher SNR due to longer dwell times, and the atom intensities are fairly similar, meaning many of the identification challenges

described in Chapter 4 do not apply here. Although the possibility of false positives and false negatives cannot be ignored due to noise and a non-uniform background, the LoG filter provides a simple, objective method for particle detection. Subsequent fitting of a 2D Gaussian to a small window around each peak refines the position with typically sub-pixel accuracy, and the use of information criteria for selecting the most likely model (Chapter 4) helps keep false detections to a minimum.

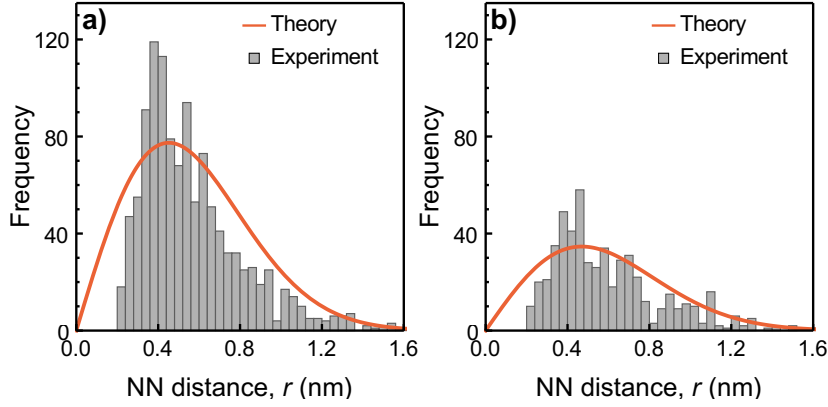
A total of 1690 atoms were extracted from Figure 6.3, giving densities  $\lambda$  of 0.78 and 0.73 nm<sup>-2</sup> respectively. Overlaying these points onto the segmented image in Figure 6.4b shows that the Cu atoms tend to bind preferentially to disordered, oxygen-rich regions, as was observed in Chapter 5 and confirmed by the DFT simulations of Figure 5.3. Characterizing this preferential binding further should give a clearer picture of how the Cu atoms interact with both the substrate and each other and, crucially, provide an insight into the effects of oxygen, which is not atomically resolved in these STEM images but evidently plays an important role.

The first step in understanding the distribution of Cu adatoms is to determine quantitatively if the atoms are distributed randomly over the substrate, or if there is a degree of clustering or dispersion not identified by qualitative inspection. Points randomly distributed on a 2D plane (or in this case a 2D projection of the effectively flat GO substrate) can be described by a spatial Poisson process, under what is known as the Complete Spatial Randomness (CSR) hypothesis [267]. If the nearest-neighbour distance between two points is  $r$ , and the point density is  $\lambda$ , then the PDF  $\phi(r)$  is a Rayleigh distribution:

$$\phi(r) = 2\pi\lambda r \exp(-\pi\lambda r^2) \quad (6.1)$$

$$\langle r \rangle = \frac{1}{2\sqrt{\lambda}} \quad (6.2)$$

Figure 6.5 shows the distribution of experimental nearest-neighbour distances for the two ADF-STEM images, and compares them to the theoretical behaviour of  $\phi(r)$  for the given atom densities. The experimental distributions match the theoretical Rayleigh distributions fairly well, suggesting that there is some degree of randomness to the Cu atom positions, with mean nearest-neighbour distances of 4.4 and 4.2 Å respectively. No significant atom clustering is apparent, at odds with the original deposition of Cu<sub>13</sub> clusters. Similar behaviour was observed in the distribution of metal atoms on graphitic carbon nitride [262], which is another promising substrate for single atom heterogeneous catalysts due to straightforward fabrication and well-isolated pinning sites at nitrogen lone pairs [23, 31]. A critical distinction is that the GO in this work is effectively a 2D plane, whereas the carbon nitride is in reality a 3D substrate that may need to be accounted for in subsequent analysis.



**Figure 6.5:** Distribution of nearest-neighbour (NN) distances for Cu adatoms on GO, extracted from Figure 6.3a and b respectively. The orange lines are the theoretical Rayleigh distributions,  $\phi(r)$ , under the CSR hypothesis for the experimental particle densities, with means (a) 4.4 Å and (b) 4.2 Å.

Despite the fact that the GO is a 2D substrate, this nearest-neighbour analysis does not take into account (a) that the minimum distance between adatoms is non-zero, and (b) that the atoms are not uniformly distributed over the plane. For example, there is qualitative evidence of clustering at relatively large length scales, around the disordered regions of GO, but this is not captured by the Rayleigh distribution analysis.

To investigate the distribution of adatoms at different length scales, a useful tool is Ripley's K-function [268, 269], which improves on the nearest-neighbour analysis by describing the characteristics of the point process at different length scales,  $h$ . It is defined as:

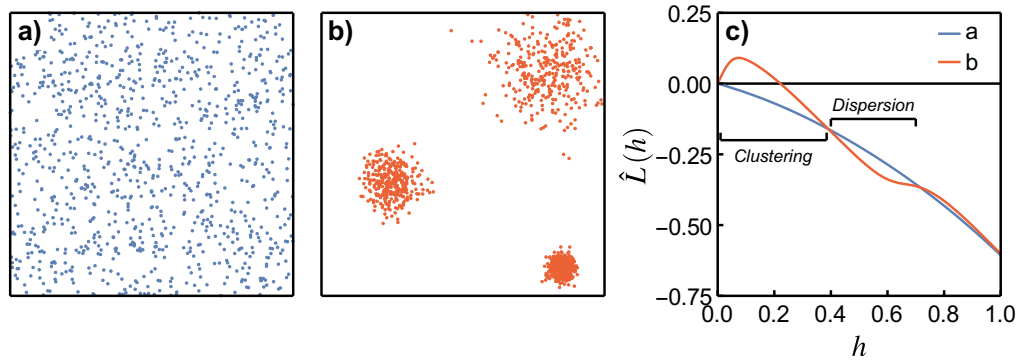
$$K(h) = \frac{1}{\lambda} \mathbb{E} \{ \text{number of additional events within distance } h \text{ of an arbitrary event} \} \quad (6.3)$$

The shape of  $K(h)$  can be used to compare experimental point sets with any type of spatial point process, rather than just the Poisson point process. If  $d_{ij}$  is the Euclidean distance between two events  $i$  and  $j$ , and:

$$I_h(d_{ij}) = \begin{cases} 1, & d_{ij} \leq h \\ 0, & d_{ij} > h \end{cases} \quad (6.4)$$

then the estimator for the K-function for  $n$  points is:

$$\hat{K}(h) = \frac{1}{\lambda n} \sum_{i=1}^n \sum_{j \neq i}^n I_h(d_{ij}) \quad (6.5)$$



**Figure 6.6:** Testing the CSR hypothesis. (a)  $10^3$  points drawn uniformly over the unit square. (b)  $10^3$  points drawn from a three-component Gaussian mixture distribution. The density,  $N_{\text{points}}/\text{area}$ , is the same in both figures. (c) Estimators of Ripley's L-function,  $\hat{L}(h)$ , for the two point sets.

Typically, the question of interest involves comparing the behaviour of  $\hat{K}(h)$  with the structure expected under a given point process. For a spatial Poisson process,  $K(h) = \pi h^2$ , and any deviations from this value are therefore indicative of either clustering or dispersion at a scale  $h$  [269]. It is useful to define the associated L-function as the standardized K-function:

$$L(h) = \sqrt{\frac{K(h)}{\pi}} - h \quad (6.6)$$

since for a set of points exhibiting CSR,  $L(h) = 0$  for all  $h$ . Applying this test to the experimental atom positions,  $\hat{L}(h) > 0$  will indicate some degree of clustering at scale  $h$ , and  $\hat{L}(h) < 0$  suggests some dispersion at scale  $h$  relative to that expected under the CSR hypothesis. To demonstrate this effect, Figure 6.6 compares  $\hat{L}(h)$  for two point processes. The first process draws points from a uniform distribution over the unit square (Fig. 6.6a), while the second draws points from a mixture of three Gaussian distributions (Fig. 6.6b). Inspecting the behaviour of  $\hat{L}(h)$  in Figure 6.6c reveals the existence of a downward bias in the case of CSR. This arises due to edge effects, since the expected number of events within distance  $h$  will be reduced for points near the boundary of the unit square. While a correction can be made for this bias [268], it is often more straightforward to make a direct comparison between the observed point sets. This shows that the points drawn from the mixture distribution are typically clustered at small scales (i.e. around the means of the Gaussian components) compared to CSR, while the dispersion at longer length scales (into the three clusters) is also highlighted.

The edge bias can also be mitigated using stochastic simulations of point processes that include the boundary regions. This approach can be used with any given point process, and thus incorporate further parameters such as non-uniformities in the PDF from which the points are drawn, without the need for a simple closed-form solution

for  $K(h)$ . By repeatedly sampling a given number of points in a region and calculating  $\hat{L}(h)$ , it is possible to obtain a measure of significance for any clustering or dispersion. Defining  $m(\hat{L}(h))$  as the number of simulations where  $\hat{L}(h)$  exceeds the experimental value, the confidence level for clustering is determined by [269]:

$$\hat{p}_{\text{clustered}}(h) = \frac{m(\hat{L}(h)) + 1}{N + 1} \quad (6.7)$$

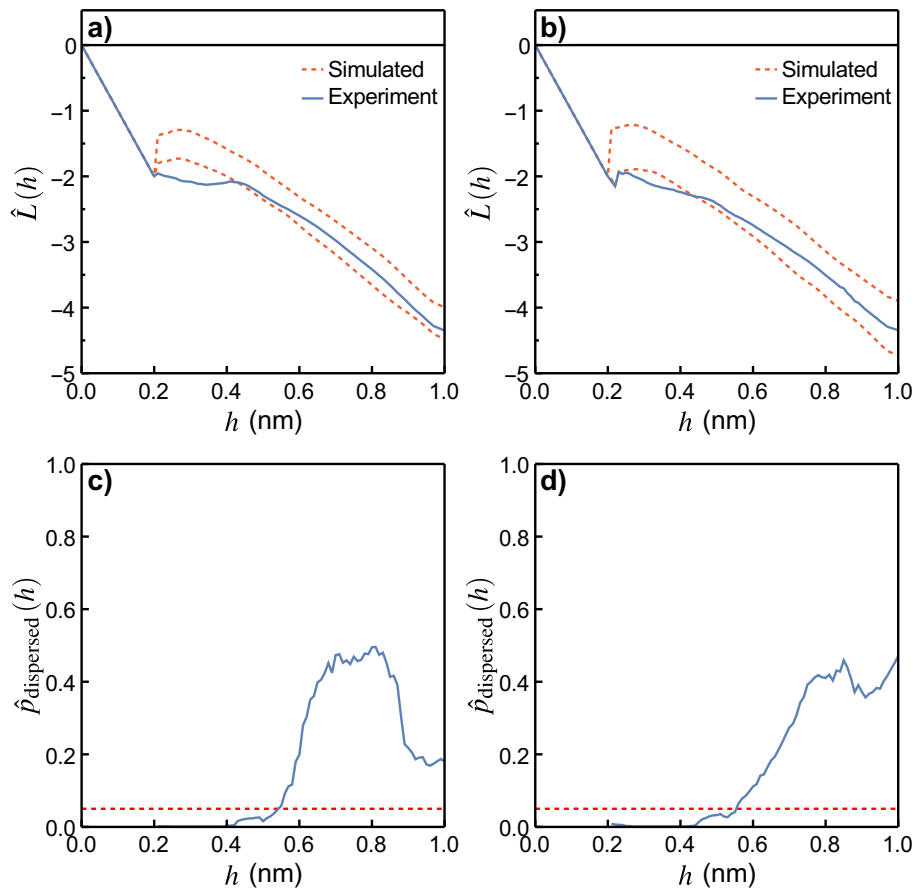
where  $N$  is the number of simulations. A similar expression can be defined for measuring the significance of any dispersion if  $m(\hat{L}(h))$  is instead taken to be the number of simulations where  $\hat{L}(h)$  is less than the experimental value.

A particular advantage of using stochastic simulations over closed-form solutions is that the minimum separation distance of the adatoms can now be incorporated. Assuming that the Cu atoms largely lie in the plane of the GO substrate, a sensible physical constraint is a minimum interatomic spacing of ca. 2 Å. This accounts for the bulk Cu–Cu distance of 2.55 Å, deviations due to STEM being a projected imaging method and the substrate not being completely flat, and finally any errors in the atomic positions (estimated to be  $\pm 0.5$  pixel = 0.2 Å).

We are now in a position to assess the degree of clustering or dispersion of the Cu atoms in a manner that closely reflects the actual distribution of adatoms on GO. To achieve this, the minimum distance constraint is combined with the sampling of random points from the disordered regions extracted in Figure 6.4b.

The top row of Figure 6.7 shows the experimental L-functions for the two GO images in Figure 6.3, taking into account the area of the disordered regions when calculating the atom density. The dashed red curves indicate the maximum and minimum values of  $\hat{L}(h)$  from  $10^3$  simulations, each sampling points in the disordered GO regions subject to the minimum separation of 2 Å. It is apparent that at large length scales the experimental atom distribution closely follows the random process, but below 5 Å there appears to be a small degree of dispersion present in the experiment not accounted for by the model. This is confirmed in the bottom row, which plots the behaviour of  $\hat{p}_{\text{dispersed}}(h)$  with the dashed line indicating a significance level of 0.05. Importantly, the shape of the experimental L-functions at small  $h$  cannot be replicated by changing the minimum separation constraint or by altering the area from which points are drawn (i.e. by changing the threshold used to segment the images in Figure 6.4b).

Combined with the nearest-neighbour study, this analysis indicates that there is a tendency for Cu atoms on GO to sit further apart than the bulk Cu–Cu nearest-neighbour distance, confirming the hypothesis that direct bonding between Cu atoms is unlikely. These observations are an important result, since the sample was initially prepared through the deposition of size-selected  $\text{Cu}_{13}$  clusters from the gas phase, yet very



**Figure 6.7:** Measuring the spatial distribution of Cu adatoms on GO. **(a,b)** Comparison of the experimental L-functions for the two ADF-STEM images with simulations incorporating the disordered structure of the GO substrate and a minimum separation constraint. The red dashed lines represent the upper and lower bounds of  $10^3$  simulations. **(c,d)** Plots of  $\hat{p}_{\text{dispersed}}(h)$  for the same simulations. The red dashed lines highlight  $p = 0.05$ .

few (if any) examples of Cu clusters are observed experimentally. This raises two important questions. The first concerns the stability of small Cu clusters on GO, and why the as-deposited clusters have broken apart into single atoms. The second question concerns the role of oxygen atoms, and whether the large Cu–Cu interatomic distances can be explained by the presence of oxygen in between the metal atoms, acting as an intermediary and thus stabilizing the Cu atoms on the substrate. This builds on the work of Chapter 5, which highlighted the weak van der Waals interactions between Cu and graphene alone. Since the positions of oxygen atoms are not clearly resolved in the experimental images, answering these questions calls for a systematic computational study of the behaviour of Cu on GO, building on the work of the previous chapter to incorporate multiple metal atoms.

A notable precedent for this work was published by Sohlberg et al. in 2004 [270], where ADF-STEM observations of Pt atoms on  $\gamma$ - $\text{Al}_2\text{O}_3$  revealed anomalously large Pt

spacings of ca. 3.5 Å. Subsequent DFT modelling of the crystalline substrate led to the discovery that not only do Pt atoms bind to oxygen sites on the support, but that the existence of an OH “cap” atop the trimer leads to a lengthening of the Pt–Pt bonds by reducing the electron density between the metal atoms. Although the OH group itself was not imaged directly, the combination of ADF imaging with DFT modelling meant that its presence could be inferred. The present work has a similar goal: can computational modelling be used to fill in the missing O atoms and thus explain the observed interatomic spacings? Although the lattice is resolved in Figure 6.2, the mobility and subsequent blurring of any oxygen atoms means STEM experiments alone are unlikely to reveal the reasons behind the large interatomic spacings. A combination of image analysis and computational modelling is thus necessary to further investigate the system.

## 6.4 Computational structure searching

In essence, computational searches for the stable structures of crystals, surfaces and clusters involve the global optimization of the free energy, bulk modulus or some other desired property. In the present scenario, it is the Cu binding energy which is of most interest, as this will dictate the overall stability of the structures seen experimentally. As was seen in Chapter 5, DFT calculations provide an effective method for calculating the binding energy of Cu atoms on GO. While classical methods using empirical potentials are considerably cheaper computationally, constructing an accurate potential for a complicated system such as Cu on GO is very difficult. First-principles methods such as DFT are thus necessary to reproduce the complex structures and properties of small metallic clusters, as well as the complex cluster–substrate interactions.

However, standard DFT approaches are only local optimizers of a structure, since the positions of the atomic nuclei are considered to be fixed during the calculation of the electronic structure (see Appendix D). To optimize the overall geometry one must calculate the forces on each atom, move them, then recalculate the electronic structure, and this is typically performed using a local gradient descent optimization algorithm. Ultimately the search for the minimum involves a non-convex function over a parameter space that scales exponentially with the number of atoms. If the initial starting points are far from the global optimum these methods can relax to a metastable local minima or fail to converge entirely, while saddle points can pose particular problems for many standard gradient-based methods [181].

Global optimization methods seek to avoid many of these problems, but are often computationally expensive and require standard local optimizers to refine the search. As a result, finding the true global minimum is often an intractable problem for



even medium-sized systems. One must also consider the “no free lunch” theorem: averaged over all possible functions, all global optimization algorithms will perform the same [271], and so no single strategy is guaranteed to give better results. Although finding an algorithm that works well in all scenarios is unlikely or even impossible, by smart application of prior knowledge and constraints the parameter space can be reduced to develop strategies for structure searching.

#### 6.4.1 *Ab initio* random structure searching

The *ab initio* random structure searching (AIRSS) technique was introduced by Pickard and Needs [231, 232], and in its simplest form places atoms randomly inside a box of random shape and/or size and relaxes the system with DFT. Rather than a random search of the full parameter space, AIRSS instead incorporates chemical knowledge such as approximate bond lengths, coordination numbers and symmetry for a given system, as well as any experimental data. Crucially, the fact that at ambient pressures atoms will not exist on top of each other means that there are very few local minima in the energy surface at high energies, which can rule out a significant fraction of the search space. Practically, the algorithm requires very few parameters while being straightforward to combine with prior experimental information.

#### 6.4.2 Alternative methods for structure prediction

Although AIRSS is the method of choice for most of this thesis, it is important to consider other successful algorithms for computational structure searching, especially in recognition of the no free lunch theorem. A brief summary is given here; for an in-depth review the reader is referred to Ref. 272.

A commonly-used Monte Carlo (MC) method is simulated annealing, which is analogous to removing defects in metals by heat treatment. The current structure is replaced by a randomly-selected nearby structure with a probability of one if it is lower in energy, and  $\exp(-\Delta E/T)$  otherwise, with  $T$  here being the annealing temperature. Simulated annealing is often a very inefficient method, and can be easily trapped in deep potential wells. Methods such as stochastic tunnelling have been developed to transform the energy surface and avoid trapping in local minima. An alternative MC approach is basin-hopping and the related minima-hopping method [273–275], which again transform the energy surface and propose random moves to overcome energy barriers between minima.

Inspired by the concept of evolution in biology, evolutionary algorithms use reproduction, mutation and recombination rules to evolve a population of structures over

several generations, using a “survival of the fittest” approach to search the energy surface [276, 277]. The USPEX evolutionary algorithm code developed by Oganov et al. is used in this chapter in the search for copper cluster structures in vacuum [278].

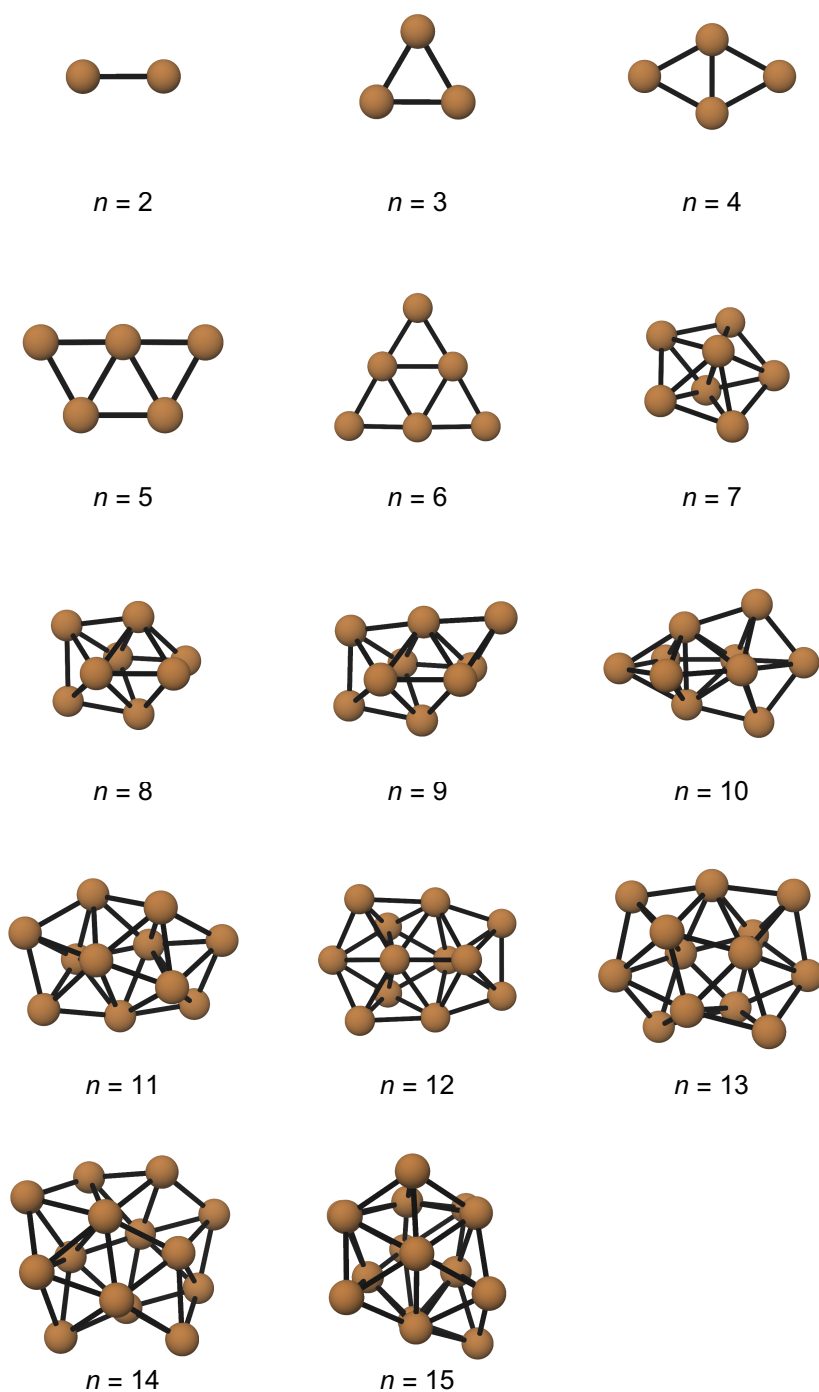
With careful consideration of the system under investigation, the optimization is not necessarily limited to methods that directly model the interactions between atoms (either classically or from first principles). Ophus et al. recently showed an elegant method for studying graphene grain boundaries using a Centroidal Voronoi Tessellation (CVT) method [279]. The CVT algorithm provides a computationally-attractive route for annealing graphene grain boundaries by treating the carbon atoms as vertices of Voronoi cells, and relaxing the lattice using Lloyd’s algorithm [280]. The relaxed tessellation will tend towards regular hexagons in two dimensions, and, when a perfect tessellation is impossible, form pentagons and heptagons in a manner very similar to defects in graphene. The relaxed structure can then be used as an input for classical or DFT calculations to investigate any out-of-plane distortions. The CVT algorithm can rapidly relax low-angle grain boundary structures containing many thousands of atoms [281], in contrast to approaches such as AIRSS which are currently limited to small, high-angle boundaries [282].

## 6.5 Structures of small copper clusters

To interpret the ADF-STEM observations, we must first understand the structures of small Cu clusters in vacuum, since this is closer to how the samples were initially fabricated. Only then can the interesting discoveries of anomalous Cu–Cu distances and the notable absence of any deposited Cu<sub>13</sub> clusters be fully understood.

### 6.5.1 Copper clusters in vacuum

The USPEX evolutionary algorithm code was used in conjunction with DFT calculations to find the low-energy structures of small Cu<sub>n</sub> clusters in vacuum, ranging in size from  $n = 2$ –15 atoms. Details of the DFT calculations are given in Appendix D. The structure search was subsequently repeated using the AIRSS technique, which gave identical results for the low-energy structures. Figure 6.8 shows the lowest-energy Cu<sub>n</sub> clusters found by the structure searches. The results are largely consistent with other works on Cu clusters [283–285], although some discrepancies in relative energies between the lowest-energy and other metastable structures were noted. Given the size of the search space, differences in computational details as well as the structure search method are the most likely reasons for the observed discrepancies.



**Figure 6.8:** The ground-state structures of small  $\text{Cu}_n$  clusters in vacuum ( $n = 2$ – $15$ ) found in a computational structure search.

It is clear from Figure 6.8 that the addition of a single atom to a cluster can have a significant effect on its shape, and by extension properties (cf. Ref. 27). The transition from a 6-atom planar structure to the 7-atom 3D structure is of particular interest. A similar transition at 6/7 atoms has been observed in simulations of Ag clusters [286], while further computational evidence suggests that anionic gold clusters adopt stable planar structures up to Au<sub>12</sub> [287]. The 7-atom pentagonal bipyramid is a common observation in transition metal clusters, and can also be seen as a building block in the larger clusters (e.g.  $n = 10$ ).

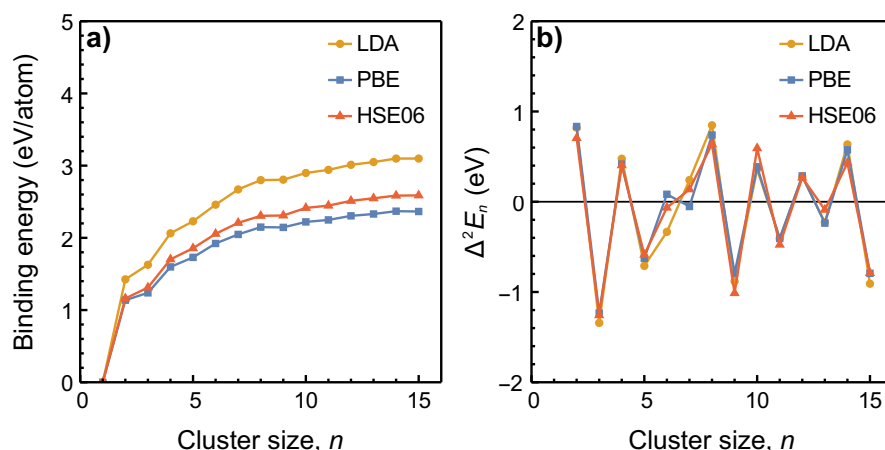
In contrast to the high-symmetry structure adopted for  $n = 7$ , a striking observation is the low-symmetry structure adopted by Cu<sub>13</sub>. For many transition metals,  $n = 13$  is a so-called “magic number”, exhibiting enhanced stability against fragmentation and adopting high-symmetry icosahedral or cuboctahedral structures [288, 289]. The argument for high-symmetry structures is typically based on a simple geometric closed-shell model or on calculations using a Lennard-Jones potential [290], but Figure 6.8 confirms that more accurate first-principles calculations are necessary to truly understand the bonding in small metal clusters, which is clearly different from the bulk behaviour. The total energy of the icosahedral Cu<sub>13</sub> cluster found in the structure search is 1.05 eV higher than the low-symmetry structure in Figure 6.8.

Since Cu<sub>13</sub> does not adopt the expected magic cluster shape, further investigation of the relative stabilities of the structures is necessary. Using the geometries shown in Figure 6.8, high-quality single-point calculations were performed using DFT to determine the free energy of the copper clusters in vacuum. As well as the generalized gradient approximation (GGA) exchange-correlation functional used in the USPEX and AIRSS searches, the lowest-energy structures were re-relaxed using norm-conserving pseudopotentials and then single-point calculations performed using the HSE06 hybrid functional. Geometry optimizations and single-point calculations were also carried out using the local density approximation (LDA) for comparison (see Appendix D for further details on the functionals).

Figure 6.9a plots the binding energy per atom as a function of the number of atoms in the cluster. All three functionals exhibit the same trend in binding energy, but this analysis does not give any significant insight into the effect of adding or removing a single copper atom. One way of quantifying the relative stability of a cluster is to consider the second-order energy differences:

$$\Delta^2 E_n = E_{n-1} + E_{n+1} - 2E_n \quad (6.8)$$

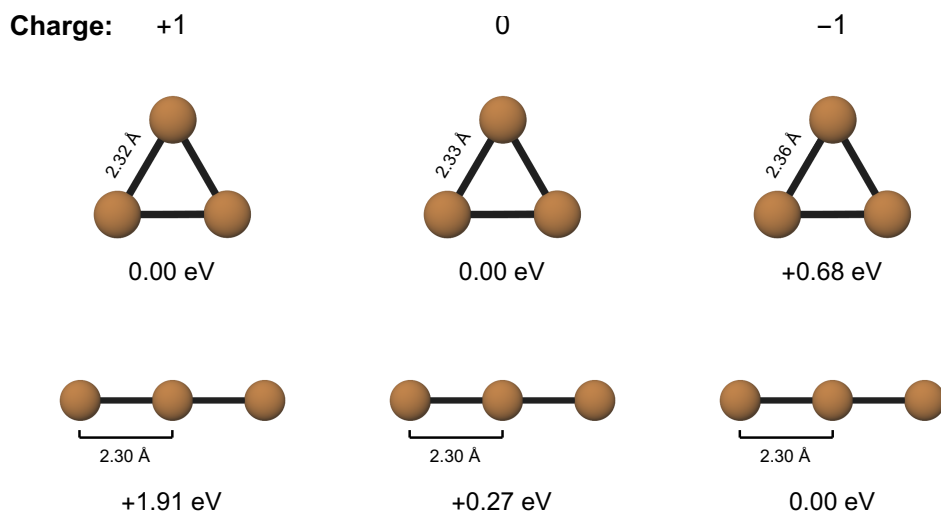
where  $E_n$  is the free energy of the Cu <sub>$n$</sub>  cluster. When  $\Delta^2 E_n \geq 0$ , this represents a local minimum in the energy of the cluster and is therefore indicative of relative stability. Figure 6.9b plots the second-order differences of the geometries shown in



**Figure 6.9:** The stability of small  $\text{Cu}_n$  clusters. **(a)** The binding energy of the ground-state structures for three different exchange-correlation functionals. **(b)** The second-order differences of small  $\text{Cu}_n$  clusters (Eq. 6.8).

Figure 6.8. It is immediately apparent that even-numbered clusters are locally stable in vacuum, whereas odd-numbered clusters (including  $\text{Cu}_{13}$ ) are unstable and thus might be expected to break apart into smaller fragments. Returning to the fabrication of the samples analyzed in this chapter, this indicates that even if the soft landing procedure did not lead to any break-up, the inherent instability of  $\text{Cu}_{13}$  suggests fragmentation at ambient temperatures over time is to be expected.

The relative stabilities calculated here are in general agreement with the comprehensive analysis performed by Itoh et al. [283] using the PW91 GGA functional. However, the planar–3D transition is again of interest, because calculations using the PBE functional predict a planar 6-atom cluster to be locally stable, while both LDA and HSE06 suggest that the 7-atom pentagonal bipyramid is stable. This is at odds with the even-numbered trend seen for all other values of  $n$ . The discrepancy arises because a structure search using the local density approximation finds the lowest-energy  $\text{Cu}_6$  structure to be 3D rather than planar. LDA is known to overbind many structures [291], and this is likely to lead to a preference for tightly-bound 3D geometries. The hybrid HSE06 functional falls in between the LDA and PBE energy values, but was not used for geometry searches due to the expensive calculations involved. Hybrid functionals such as HSE06 are often employed to give “better” results, particularly in terms of energy or geometry. However, it was recently reported that the flexibility of parameterized hybrid functionals can actually come at the expense of an accurate representation of the true electron density [292]. While this may not matter greatly for calculations involving bulk systems, for scenarios such as the small clusters considered here it may be very important, and again highlights the need for studies comparing experimental observations with DFT predictions.



**Figure 6.10:** The effects of charge on the ground-state structure of  $\text{Cu}_3$  clusters in vacuum. From left to right, the structures of a cationic (+1), neutral (0) and anionic (-1) cluster. Relative energies above the minimum are given, and Cu–Cu bond lengths are also shown.

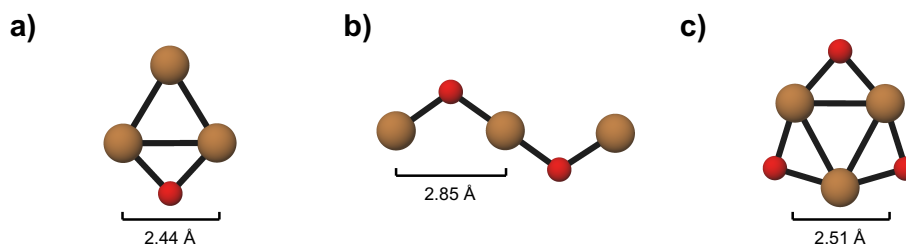
### 6.5.2 Charged copper trimers

The  $\text{Cu}_3$  cluster in Figure 6.8 adopts an equilateral triangle configuration, with Cu–Cu spacings of 2.33 Å. This contrasts with both the spacings and the asymmetry of the  $\text{Cu}_3$  motifs seen in the STEM images (Fig. 6.2), and indicates that ground-state DFT calculations of isolated clusters are insufficient to explain the experimental observations.

Before investigating the effects of the GO substrate, which is known to be a complex and highly disordered material, it is worth conducting further inexpensive calculations on the isolated clusters to learn more about the structure and the effects of properties such as charge. Electrical charge can have a significant effect on the structure of some metal clusters (for example the stabilization of the planar structure of  $\text{Au}_{12}$  [287]), so it is the first parameter investigated here. Figure 6.10 shows the two stable structures of a  $\text{Cu}_3$  cluster in vacuum calculated using the PBE functional, with relative energies also shown. While the cationic and neutral clusters prefer the equilateral triangle motif, the anionic cluster favours a linear configuration [293]. However, all the structures exhibit a typical Cu–Cu bond length of ca. 2.3 Å and none of the asymmetry seen experimentally, so charge alone cannot account for the anomalous Cu–Cu distances.

### 6.5.3 Oxidized copper trimers

The next step is to incorporate the presence of oxygen into the structure searches, since the experimental clusters are in fact deposited on graphene oxide, and we know



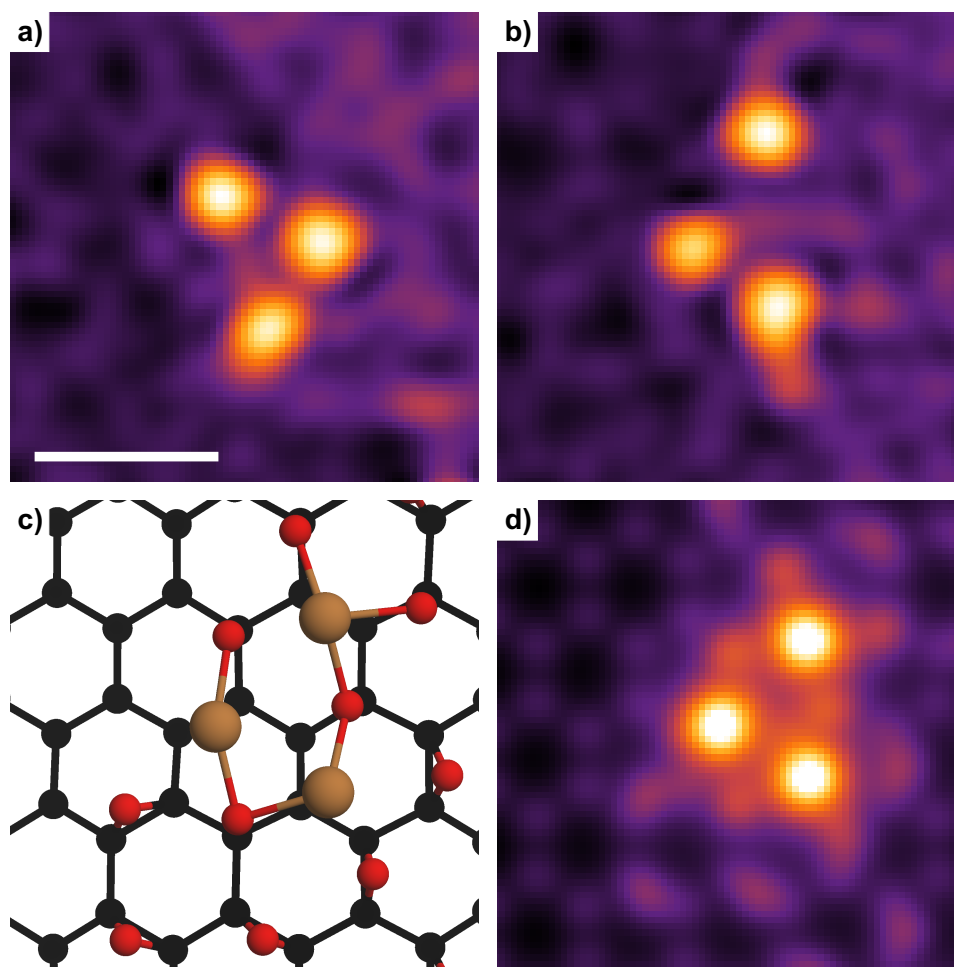
**Figure 6.11:** The effect of oxygen on the ground-state structure of  $\text{Cu}_3\text{O}_x$  clusters in vacuum for  $x = 1\text{--}3$  O atoms (in red). Cu–Cu interatomic distances are highlighted.

that Cu binds preferentially to oxidized regions of GO. Furthermore, the clusters and the substrate were exposed to air prior to being inserted into the microscope. Before considering the full system incorporating the graphene oxide substrate, the structure of an isolated, oxidized  $\text{Cu}_3$  trimer is briefly considered due to the significantly lower computational cost (see Ref. 294 for extensive oxidation studies of other copper clusters). Using the AIRSS approach, systems consisting of 3 Cu atoms and 1–3 O atoms were relaxed. The lowest-energy structures are given in Figure 6.11. While the Cu–Cu interatomic distances are now larger than the pure copper clusters, the increase in size is still not sufficient to fully explain the ADF-STEM observations. That said, the linear configuration of  $\text{Cu}_3\text{O}_2$  is particularly interesting (Fig. 6.11b), since the oxygen atom acts as an intermediary between neighbouring copper atoms, pushing them apart to the extent that they are no longer directly bonded.

#### 6.5.4 Copper trimers on graphene oxide

The results above indicate we must consider the complicated environment of the substrate to fully understand the observed  $\text{Cu}_3$  motif. Using a similar approach to the single copper atom simulations of Chapter 5, extensive AIRSS searches were carried out, incorporating a loose  $\text{Cu}_3$  trimer “unit”. This acts as an additional constraint and helps to bias the search towards triangles rather than lines of atoms.

As observed in Chapter 5, the interaction between a Cu atom and pristine graphene is very weak, arising as a result of van der Waals interactions only. Accordingly, searches involving  $\text{Cu}_3$  atop a pristine graphene sheet resulted in small equilateral triangles akin to the isolated cluster. Further searches involving vacancies or pores in graphene were also inconsistent with the experimental observations. The regularity of the graphene pores typically led to Cu arrangements that were too large to match the STEM observations. Furthermore, a Cu cluster sitting in a graphene pore is ruled out by the high-resolution STEM data, which indicates that the trimers sit favourably at boundaries between  $\text{sp}^2$  and  $\text{sp}^3$  regions on graphene oxide. Building on this,

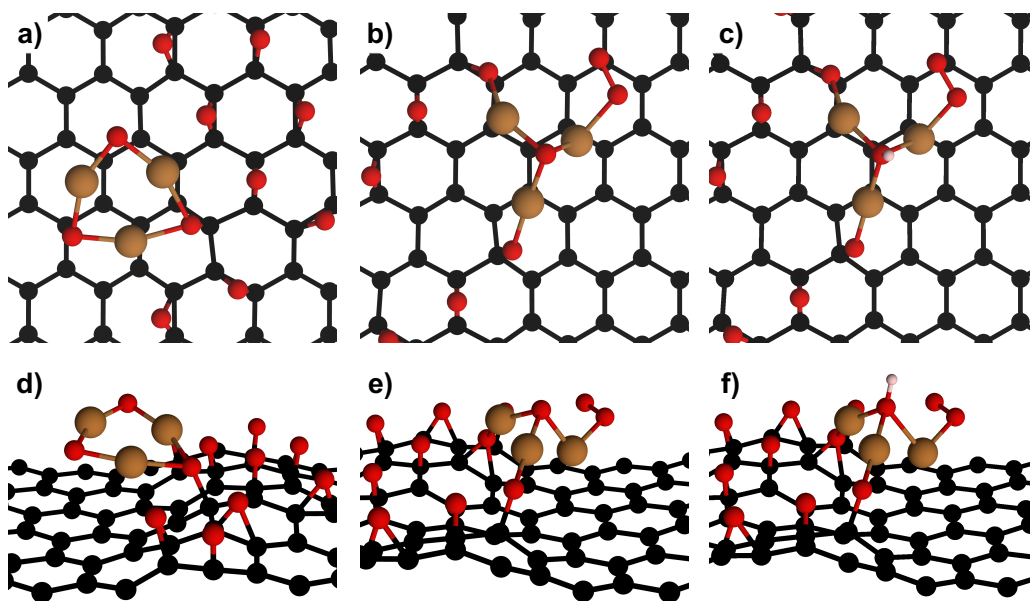


**Figure 6.12:** (a,b) Two Cu<sub>3</sub> trimer motifs with anomalous Cu–Cu interatomic spacings taken from the insets of Figure 6.2. Scale bar: 0.5 nm. (c) One of the possible low-energy structures found using AIRSS. Oxygen atoms are red, carbon atoms are black. (d) A simulated ADF-STEM image of the structure in c. Multislice parameters were selected to match the experimental setup.

further AIRSS studies were performed including small numbers of oxygen atoms (with randomized positions) alongside the Cu<sub>3</sub> unit.

While many different structures were found in the searches, no definitive picture emerged that fully explains the asymmetric trimers seen experimentally. This is highlighted by Figure 6.12c, which shows one of the low-energy structures found with AIRSS. The multislice simulation in Figure 6.12d shows a remarkable similarity to the two ADF-STEM images (Figs. 6.12a and b) in terms of size, asymmetry and the intensity in between Cu atoms attributed to oxygen. The argument that preferential binding to oxygen causes an increase in the Cu–Cu spacing is seemingly convincing, but the structure in Figure 6.12c is not in fact the lowest-energy structure found using AIRSS. It actually lies ca. 2 eV above another structure (Fig. 6.13b) that consists of a small equilateral triangle of Cu atoms with an oxygen “cap”, akin to the Pt





**Figure 6.13:** (a) The  $\text{Cu}_3\text{O}_3$  cluster bound to graphene oxide, as found using an AIRSS search. (b) The lowest-energy structure found from an AIRSS search involving 3 Cu atoms on graphene oxide. (c) Replacing the oxygen cap in b with an OH group causes the Cu–Cu bonds to lengthen slightly. (d–f) Side-on views of the three structures highlight the buckling of the graphene sheet caused by oxygen functional groups.

trimer analyzed by Sohlberg et al. [270]. Figure 6.13a meanwhile shows how the  $\text{Cu}_3\text{O}_3$  cluster found in an isolated structure search (Fig. 6.11c) can also bind to the substrate, and this structure is ca. 1.5 eV lower than the asymmetric trimer identified in Figure 6.12c.

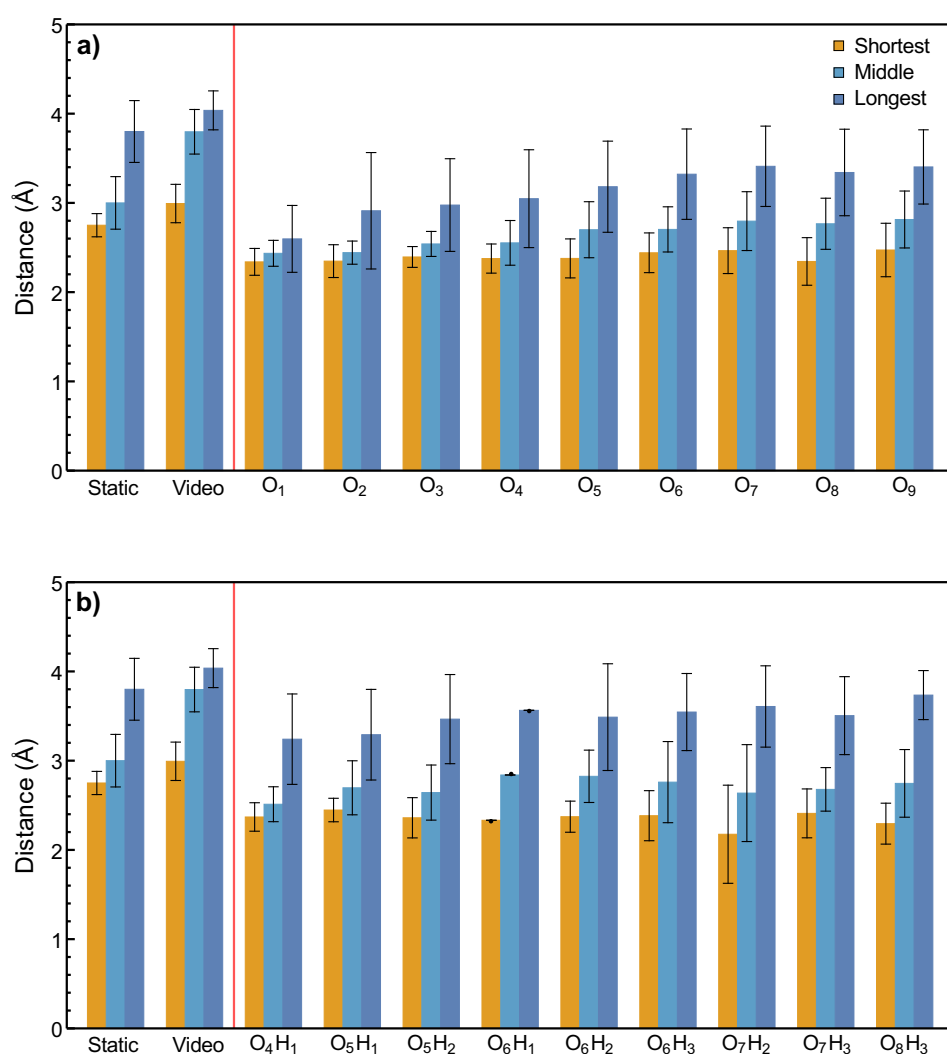
An interesting finding from Ref. 270 was the notion that an OH group, rather than an oxygen atom, can cause the metal–metal bond to lengthen as a result of transfer of electrons from the bond to the electronegative OH cap. This effect is tested in Figure 6.13c, where the oxygen cap has been replaced by an OH group and re-relaxed with DFT. While the symmetry is not altered by the cap, there is a marginal increase in the Cu–Cu distances, from an average of 2.65 Å to 2.80 Å. Hydrogen atoms and OH groups are certainly present in the experimental system, with GO known to contain both epoxy and hydroxyl groups [217]. Although the increase in Cu–Cu distances are again not sufficient to explain the STEM observations, Figure 6.13c does highlight how important it is to consider the effects of functional groups on a complex substrate such as GO.

These results highlight an issue with the AIRSS calculations performed here, in that the size and complexity of the system (ca. 100 atoms including the graphene sheet) makes an exploration of the full energy landscape extremely difficult. Figure 6.12c may represent a possible  $\text{Cu}_3$  structure, but it is not possible to say how likely it is to

form compared to other, lower-energy structures such as Figure 6.13. What can be drawn from Figure 6.12c is that oxygen between Cu atoms do indeed a cause of the lengthening of interatomic distances, building on the isolated cluster results, as well as stabilize the pinning of Cu to the substrate. Ultimately, it would seem that the AIRSS searches as currently designed are perhaps unlikely to locate the exact structures of the Cu<sub>3</sub> trimers seen experimentally, because the locations and number of oxygen and hydrogen atoms around the copper are essentially unknown quantities. This is without considering the fact that DFT is a ground-state method and that Figure 6.1 shows the Cu atoms to be mobile under the electron beam and therefore likely to exist in an excited state. The possible role of contaminants such as Si (identified in Chapter 5) have also been ignored here. Rather than run further random searches that may never find the desired structure, AIRSS can instead be used to provide valuable insights such as the effect of intermediary oxygen atoms in both the lowest-energy structures and in other metastable structures. Such insights may then prove useful for the design of future experiments to finally pin down the behaviour of small transition metal clusters on both GO and other disordered or poorly-understood substrates.

To that end, further searches were conducted to provide a more systematic study of the effects of both O and H atoms on the Cu–Cu interatomic distances of a Cu<sub>3</sub> motif on GO. All structures within 3 eV of the lowest-energy structure were considered for this analysis. Figure 6.14a shows how increasing the number of oxygen atoms in the system leads to a corresponding increase in the length of the longest Cu–Cu spacing. Note that all distances are in projection, to mimic the experimental conditions of an ADF-STEM image. However, the shortest Cu–Cu spacing remains fairly constant, and is consistently shorter than the experimental observations. Figure 6.14b shows a similar trend when OH groups are introduced to the model and, in keeping with the idea from Ref. 270, the spacings are generally larger than the oxygen-only case.

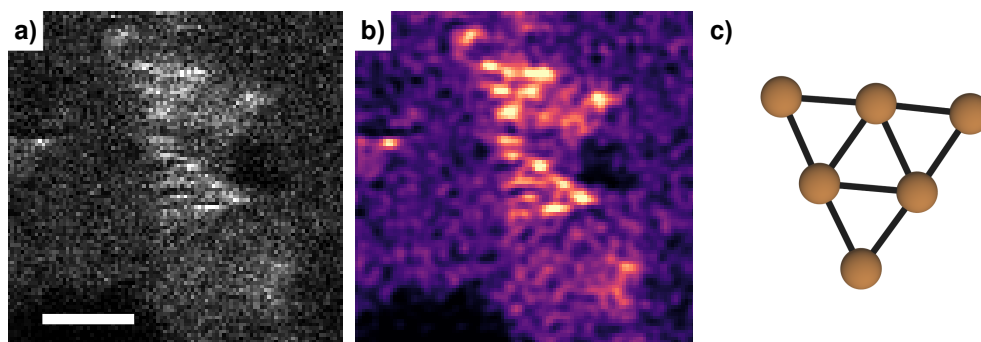
The DFT analysis in Chapter 5 indicated that a strong binding between a Cu adatom and the substrate required at least two oxygen atoms per Cu atom; AIRSS searches with a single Cu and a single O atom repeatedly led to Cu–O motifs detached from the graphene substrate. If a binding energy of ca. 3 eV per atom is required for reliable ADF imaging, it is reasonable to assume that the same Cu:O ratio applies to the larger motifs including the trimer. Figure 6.14 therefore represents a study with only the bare minimum of oxygen and hydrogen atoms, and increasing the number even further is a reasonable next step. One might expect that surrounding each Cu atom in the trimer motif with two or more O atoms may force the Cu atoms to sit at least 3 Å apart, in keeping with the spatial distribution analysis performed in Section 6.3. Such an arrangement would then be stable enough to image, in contrast to the smaller equilateral trimers, where the Cu binding energy is only 1.1 eV (Fig. 6.9a). Increasing the oxygen (and hydrogen) content around the trimer motif is an avenue for future study.



**Figure 6.14:** Comparing the Cu–Cu interatomic distances in the  $xy$ -plane (parallel to the graphene sheet) of the  $\text{Cu}_3$  motif investigated with AIRSS. The experimental distances from the static images (Fig. 6.2) and the video (Fig. 6.1) are also shown. **(a)** Distances from simulations involving oxygen atoms only. **(b)** Distances from simulations involving both oxygen and hydrogen atoms. The colours represent the shortest, middle and longest sides of the triangle motif (see legend).

### 6.5.5 Larger copper clusters on graphene oxide

Along with the single Cu atoms and the  $\text{Cu}_3$  motif discussed previously, a small number of larger clusters were observed on the GO substrate. Figure 6.15 shows one such example of a  $\text{Cu}_6$  cluster taken from a region of Figure 6.3a. Unlike the smaller 3-atom structures seen before, this image matches the planar structure found in the structure search of isolated clusters, with each side of the triangle measured ca. 5.2 Å. This suggests that GO substrate has less of an influence on the cluster than in the  $\text{Cu}_3$  case, perhaps stemming from the results in Figure 6.9b showing that the even-numbered

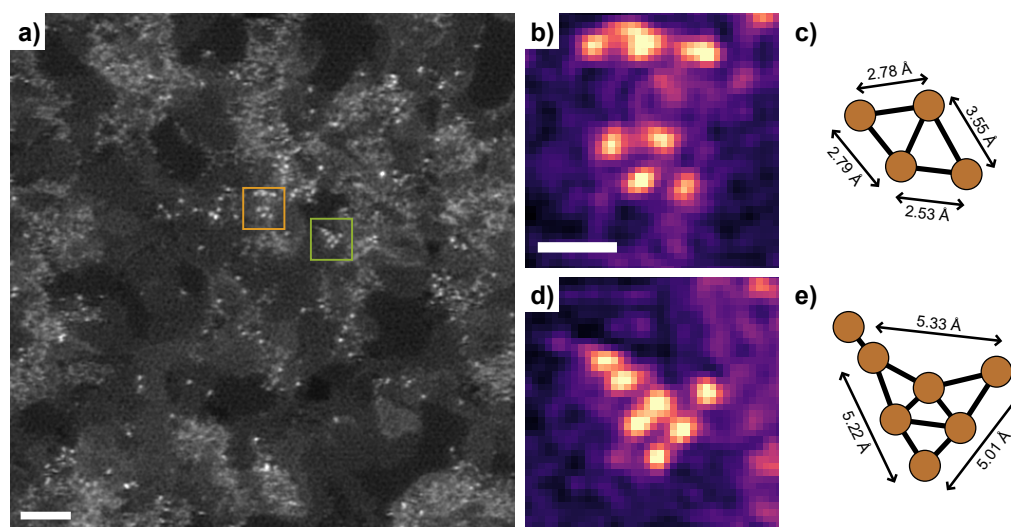


**Figure 6.15:** (a) Unfiltered ADF-STEM image of a  $\text{Cu}_6$  cluster on GO, extracted from Figure 6.3. Scale bar: 1 nm. (b) The image after Gaussian filtering. (c) The lowest-energy  $\text{Cu}_6$  cluster in vacuum found using a computational structure search (see Fig. 6.8).

clusters are locally stable. Another point to consider is that the binding energy per atom in the  $\text{Cu}_6$  cluster is nearly twice as strong as the  $\text{Cu}_3$  trimer (Fig. 6.9a).

Further evidence of larger Cu clusters is presented in Figure 6.16. The first highlighted region shows a  $\text{Cu}_4$  tetramer similar to the isolated cluster found with AIRSS, albeit with one significantly larger interatomic spacing (Fig. 6.16c). The shape suggests that the role of oxygen and other functional groups on the substrate may be less significant for the even-numbered cluster than for the  $\text{Cu}_3$  motif, linking back to the stability trends in Figure 6.9. However, the substrate effects are certainly not negligible in the  $\text{Cu}_4$  case, but without higher-resolution STEM data it is difficult to identify whether there are O atoms sited between the Cu adatoms.

The cluster shown in Figure 6.16d is particularly interesting, since it features the triangular motif of the planar  $\text{Cu}_6$  cluster with comparable interatomic spacings. In the top-left corner there appears to be an additional Cu atom, with two possible explanations to consider. The most likely reason given the short distance ( $1.33 \text{ \AA}$ ) between this atom and its neighbour is that they may in fact be the same atom. Beam-induced motion has thus led to it being imaged twice as the probe scans across the sample. The alternative explanation is that the structure is indeed a  $\text{Cu}_7$  cluster, which then raises the question of how the GO substrate stabilizes the planar geometry over the pentagonal bipyramid of the isolated cluster (Fig. 6.8), and whether this stabilization mechanism also applies to larger cluster sizes. DFT calculations could provide the answer, but without better characterization of the oxygen atoms and the substrate, any study will likely be larger and more complex than the  $\text{Cu}_3$  scenario presented previously. However, the orientation of the planar  $\text{Cu}_6$  unit relative to the substrate may in fact form a fairly tight constraint and reduce the scale of the search.



**Figure 6.16:** (a) Filtered ADF-STEM image of copper adatoms dispersed on graphene oxide. Scale bar: 1 nm. (b) An enlargement of the orange box highlighted in a, showing evidence of a Cu<sub>4</sub> cluster. Scale bar: 0.5 nm. (c) Cu–Cu interatomic spacings of the cluster. (d) An enlargement of the green box highlighted in a, showing evidence of a Cu<sub>6</sub> cluster, with a possible seventh atom at the top-left corner. (e) Cu–Cu interatomic spacings of the cluster.

## 6.6 Summary

The drive towards sub-nanometer clusters and spatially-dispersed single atoms for catalysis has highlighted the need for novel characterization and simulation techniques to address the complex structures seen in these heterogeneous systems. In particular, there remains a gap between theory and experiment regarding the interactions between the catalyst and the support. This chapter has demonstrated some of the ways in which aberration-corrected STEM and DFT can be combined to bridge this gap and guide the design and fabrication of active, stable catalysts on industrially relevant substrates. Quantitative studies of the spatial distribution of adatoms as presented here can be used to determine the stability of single atom catalysts in the early stages of a reaction. The time-resolved observations of the Cu<sub>3</sub> trimer motif show how this can be extended to measure the effects of agglomeration or fragmentation over time or in response to external stimuli.

Building on the quantitative STEM analysis, the second half of the chapter explored the potential applications of structure search methods for studying the structure of heterogeneous catalyst systems. In conjunction with the experimental observations, the AIRSS method showed how the structure of small transition metal clusters on a support can be very different to the structures found in vacuum. A full explanation of the effects of GO on Cu clusters remains uncertain, but the evidence here points towards a need for future computational studies to consider the substrate in greater detail. While previous work has explored the structure of clusters on idealized, crystalline substrates

such as graphene,  $\text{TiO}_2$  or  $\text{Al}_2\text{O}_3$ , as substrates such as functionalized graphene or graphitic carbon nitride become more widespread modelling must take into account the complexities inherent in these supports. The work here has shown that the functional groups intended to pin the copper to the substrate can have a significant effect on the metal clusters. AC-STEM, and in particular the combination of ADF imaging and EELS for characterization with advanced analytical tools, is well-placed to provide the necessary information for developing new computational studies. One such example is the recent use of vibrational EELS to directly detect the presence of water and related species including OH groups [295].

## Chapter 7

# Adatom dynamics and the Si(110) surface

### 7.1 Introduction

Surface dynamics lies at the heart of many areas of materials and chemical science, and that characterizing surface adsorption, reactions and diffusion at the atomic scale is crucial to understanding and controlling materials at the single atom level. This chapter applies the time-resolved STEM methods developed in this thesis to the study of crystalline substrates, namely the (110) surface of silicon\*. Whereas heterogeneous substrates such as GO are particularly relevant for catalysis, crystalline surfaces have numerous important applications such as epitaxial growth and device fabrication. The silicon surface in particular is one of the most widely-studied due to its importance in the semiconductor industry, and understanding the growth of oxide and metal films for metal–oxide–semiconductor field-effect transistors (MOSFETs) is key as these devices become ever smaller.

On the experimental front, SPM techniques are well-suited to studying clean surfaces [83], and are increasingly capable of in-situ surface studies with both gases and liquids [97], and at rates of 10 fps or higher [297]. However, the hardware challenges with in-situ SPM can be significant, and time-resolved STEM thus provides an alternative method with comparable temporal and spatial resolution, and facilitates integrated characterization through quantitative imaging and atom-by-atom spectroscopy. Recently, cross-sectional STEM has been used to explore the reconstruction of clean SrTiO<sub>3</sub> at elevated temperatures [298]. In a similar vein, the nanoparticle dynamics examined with the PGURE-SVT algorithm in Figure 3.10 revealed the reconstruction

---

\*The work presented in this chapter was performed in collaboration with D. Knez, E. Schmidt, R.K. Leary, F. Hofer, P.D. Bristowe and P.A. Midgley [296]. D.K. prepared the samples and acquired the STEM data. E.S. developed the atom detection and tracking code (see Chapter 4).

of the nanoparticle edges in projection. However, neither approach is able to resolve structural changes in the surface plane.

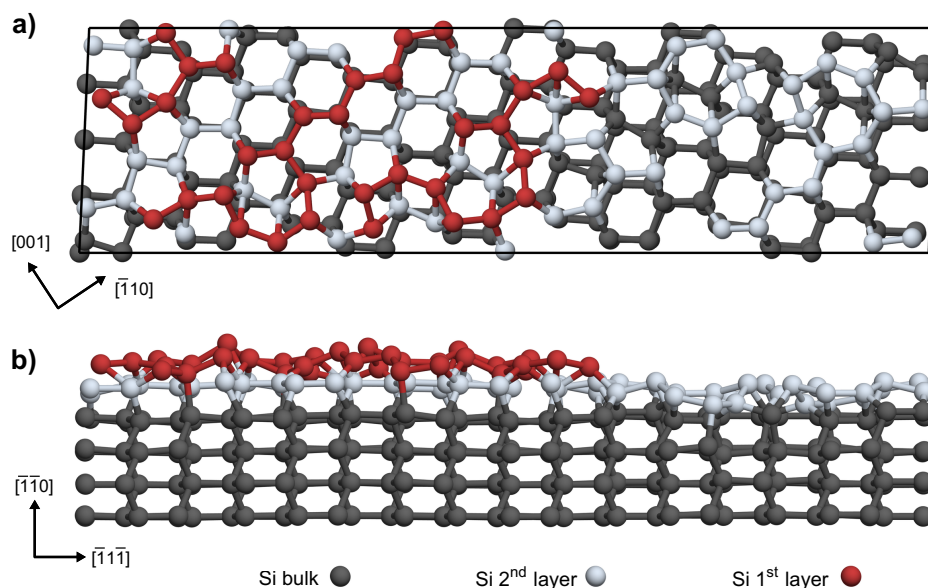
Developing new methods for recovering surface information using STEM is the first challenge addressed in this chapter, utilizing the computational imaging techniques developed in this thesis to explore the Si(110) surface dynamics with atomic resolution. Pt adatoms, deliberately sputtered onto the Si surface, make the system under investigation in this chapter particularly interesting. The presence of a crystalline substrate brings a significant benefit to the image analysis, since the periodic lattice will be, by definition, low-rank. This means the motion of any surface adatoms can be separated using the ORPCA algorithm described in Chapter 3, and studied independently of the Si surface. The first half of this chapter focuses on the surface, employing molecular dynamics (MD) simulations to explore the distinctive surface features observed in the Si substrate. The second half of the chapter is dedicated to the analysis of beam-driven diffusion of Pt adatoms. In Chapter 5, the diffusion of metal adatoms on graphene oxide was shown to be anomalous as a result of the heterogeneity present in the substrate. In contrast, the relative homogeneity of crystalline surfaces, with well-defined pinning sites, means that models of atomic diffusion are more thoroughly understood, thus providing an interesting comparison to the anomalous diffusion work described previously.

## 7.2 Silicon surface studies

Atoms at the surface of any material will experience vastly different forces and interactions compared to the bulk, resulting in a change in the equilibrium positions of atoms close to the surface. In a crystalline material this change can be a simple relaxation of the first few layers of atoms, or it can lead to a significant restructuring with changes in both atomic spacing and symmetry in order to minimize the surface energy. Interactions with adsorbates such as liquids or gases can also play a significant role in the reconstruction of a surface, particularly when the clean surface features a significant number of unfavourable dangling bonds.

Direct imaging of surface reconstructions is typically performed in ultra-high vacuum conditions using STM, which famously revealed the long-range  $7\times 7$  reconstruction of the Si(111) surface [299]. Structures can also be inferred from reflection high-energy or low-energy electron diffraction (RHEED and LEED), which were used to determine the simpler  $2\times 1$  reconstruction of Si(100) [300]. However, these methods do not recover aperiodic features in the same manner as STM. Spectroscopic techniques can provide complementary information about the surface chemistry, and the combina-





**Figure 7.1:** The  $16\times 2$  surface reconstruction of Si(110). (a) Viewed from above, surface atoms are highlighted in red, identifying the channels along the  $[\bar{1}10]$  directions. (b) A view along the  $[\bar{1}12]$  direction clearly shows the surface step detected in STM studies. Structure taken from Ref. 303.

tion of STEM imaging and EELS is particularly attractive for performing chemical characterization.

In contrast to the other low-index surfaces of Si, the clean (110) surface is known to exhibit a much larger  $16\times 2$  reconstruction, shown in Figure 7.1, with steps and terraces running in  $\langle 112 \rangle$  directions [301]. Theoretical calculations using DFT have indicated that pairs of Si pentagons arranged into zigzag chains (confirmed experimentally) are crucial to the stability of this surface [302, 303].

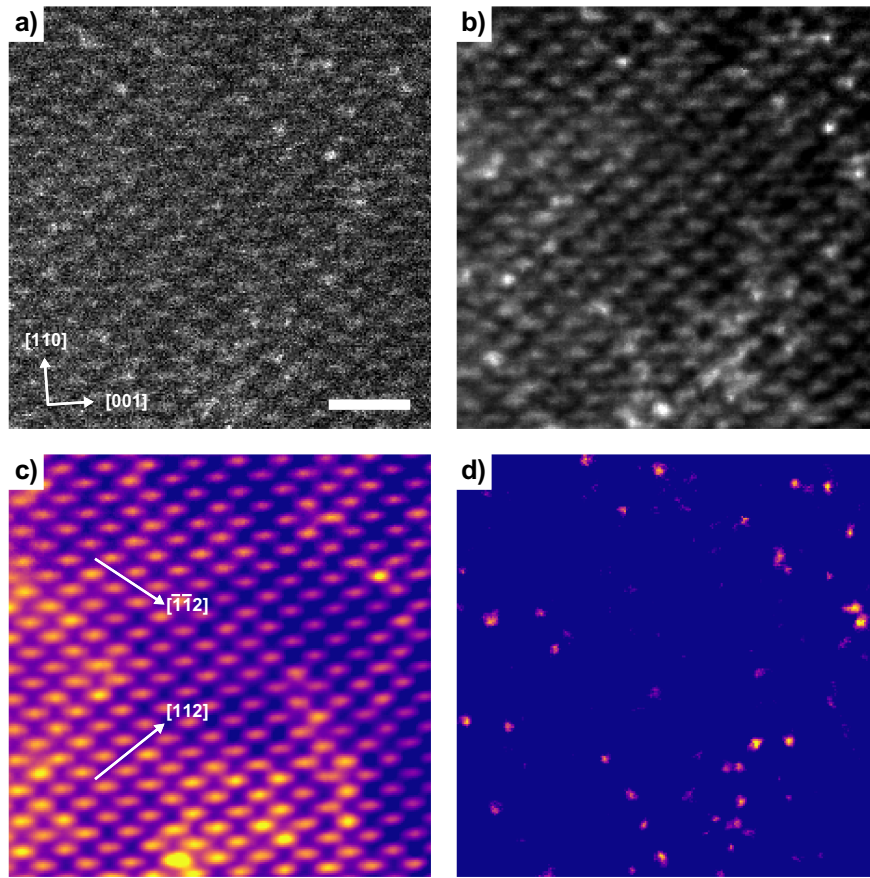
Subsequent investigations focused on the oxidation behaviour of the  $16\times 2$  reconstruction have identified preferential growth in the  $\langle 112 \rangle$  direction at these pentagonal pairs, with oxygen atoms clustering together during the early stages of oxidation [304–306]. Similar oxygen agglomeration has been observed using infrared (IR) spectroscopy and theoretical calculations on the Si(100)- $2\times 1$  reconstruction at high temperatures [307]. This clustering at the onset of oxidation can be considered an early precursor to the Si/SiO<sub>2</sub> interface [308], which was studied by Pantelides et al. using a combination of Z-contrast imaging and EELS [309]. While there is extensive literature on silicon oxides [310], there has been considerably less work on the initial stages of oxide formation, and only with high-resolution imaging techniques such as STEM or SPM and extensive computational modelling can these mechanisms be understood.

Finally, metal adatoms are known to play a key role in oxidation, and their dynamic behaviour is also critical in the initial stages of metal silicide film growth [311]. An atomic-scale understanding of the metal–Si interaction is therefore fundamental to controlling device fabrication at the atomic scale.

### 7.3 ADF-STEM results

A thin Si lamella was prepared from a monocrystalline wafer using focused ion beam (FIB) milling, using Ga ions and an accelerating voltage of 30 keV. Final thinning was performed using an Ar ion beam (900 eV energy) at an angle such that Pt atoms from the Pt–C deposit used to affix the lamella to the TEM grid were deliberately sputtered over the Si surface. Annular dark-field (ADF) STEM imaging was performed on an FEI Titan<sup>3</sup> (S)TEM operating at a primary beam energy of 300 keV. The beam convergence semi-angle was 20 mrad, and the ADF detector angle range was 60–200 mrad. The probe current was 80 pA and the pixel dwell time was 3  $\mu$ s. Time-resolved STEM imaging was performed at 4 fps, giving an estimated dose rate of  $2 \times 10^6 \text{ e}^- \text{ nm}^{-2} \text{ s}^{-1}$ . The accelerating voltage in this experiment is significantly higher than that used in Chapter 5, and so is almost certain to induce changes in both the substrate and the adatoms due to collisions between the energetic electrons and the atomic nuclei. As with the GO experiments, the STEM probe will excite the Pt adatoms out of stable binding sites, and any resulting motion is likely to be similar to that seen in the high temperature conditions typical of annealing, oxidation and film deposition.

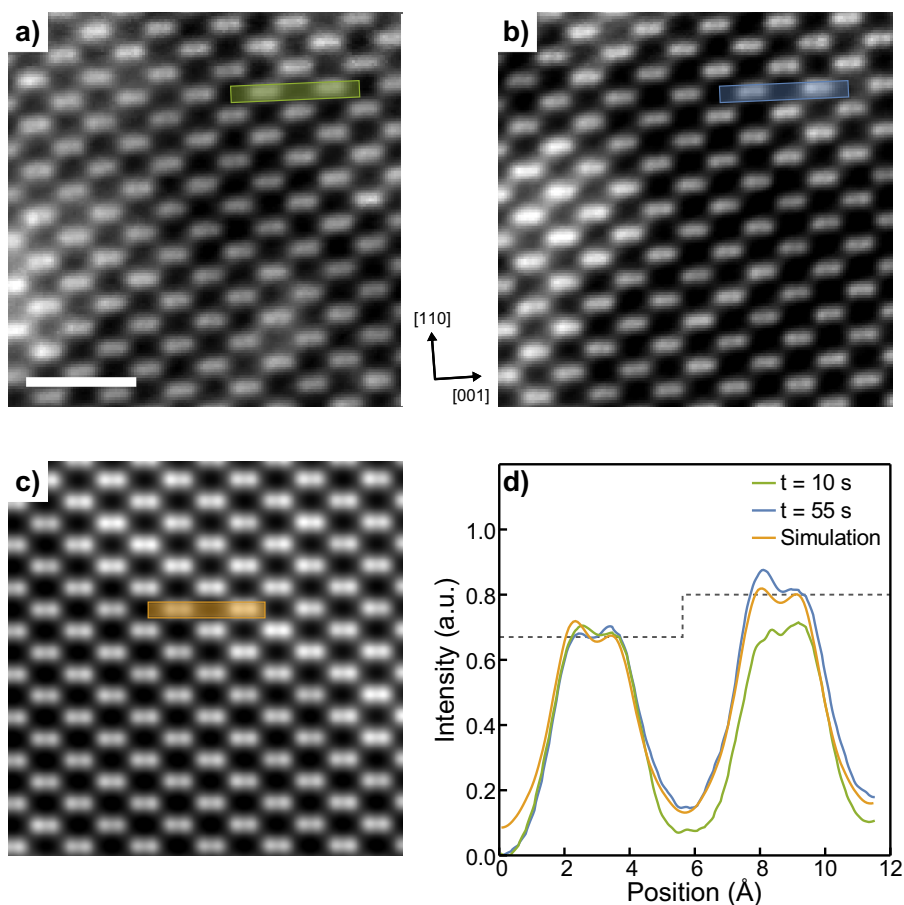
The acquired image sequences were first denoised with the PGURE-SVT algorithm, exploiting local spatio-temporal correlations. This initial step is necessary due to the presence of significant noise and sample drift in the sequences, and the denoising was found to drastically improve the performance of the subsequent analysis. Crucially, removing the sample drift through image alignment was found to be much faster and more robust using the denoised data. The alignment was performed with a non-rigid registration algorithm designed for STEM imaging [174]. Figure 7.2a shows an as-acquired frame from one of the sequences, and the denoised frame is shown in Figure 7.2b for comparison. The characteristic “dumbbells” of Si(110) are considerably clearer following noise filtering, and the contrast of the heavier Pt adatoms is enhanced relative to the Si background. Preprocessing the dataset in this fashion makes the separation of adatoms and Si lattice using ORPCA a far easier task, and the results in Figures 7.2c and d show that the foreground and background are cleanly separated into the two components. Any changes in the substrate can now be analyzed independently of the diffusive adatom behaviour. The full sequence is available on the accompanying CD (see Movie S14).



**Figure 7.2:** (a) A frame from an ADF-STEM sequence of Pt adatoms on Si(110). Scale bar: 1 nm. (b) The same frame following processing to remove noise. (c) The low-rank component recovered by ORPCA processing. (d) The sparse component containing the Pt adatoms after applying ORPCA. See also Movie S14.

## 7.4 Imaging the silicon (110) surface

Figure 7.2c exhibits several bright “chevron-like” features on the Si surface running in  $\langle 112 \rangle$  directions. These features are even clearer in a second sequence (Figs. 7.3a and b and Movie S15), where the Si atomic columns are resolved into distinct dumbbells (spacing  $1.36 \text{ \AA}$ , Fig. 7.3d). The  $\langle 112 \rangle$  direction is particularly significant, since it is both the direction of the steps on the clean  $16 \times 2$  reconstruction, and also the preferred growth direction for oxide formation on Si(110). Determining the exact origin of these features is thus important for understanding how the beam is affecting the substrate, especially as the features are seen to move and become sharper over the course of the sequence (Movies S14 and 15).



**Figure 7.3:** (a) The low-rank background of an ADF-STEM sequence of Pt adatoms on Si(110). Scale bar: 1 nm. (b) The same region 45 s later (see also Movie S15). (c) A simulated ADF-STEM image of the reconstructed surface (Fig. 7.1), with the step running diagonally across the image. (d) Line profiles from the experimental and simulated STEM images in b and c. The dashed grey line indicates the position of the step on the clean surface.

#### 7.4.1 The Si(110) surface

One of the advantages of ADF imaging discussed throughout this thesis is that the scattered intensity scales with the sample thickness and the atomic number, facilitating elemental characterization and atom counting with atomic resolution, as well as direct comparison with theoretical structures and simulated ADF images [58]. Using Z-contrast imaging, analysis of the intensity of a Pt atom atop a Si atomic column compared to a Si column only gave an estimated thickness of the lamella of 7 nm. Quantitative analysis can also be used to assess whether the bright  $\langle 112 \rangle$  features are caused by the Si surface itself or by oxidation, noting that O atoms are expected to be more mobile under the electron beam and thus be blurred during the image acquisition and processing. It is also highly likely that the electron beam will preferentially sputter off oxygen and other low atomic number contaminants [312].

ADF-STEM simulations were performed using the multislice method, accounting for the thermal diffuse scattering contribution to ADF-STEM image intensity. The images were convolved with a Lorentzian function with full-width half-maximum (FWHM) of 1.2 Å to account for the finite source size effects. Figure 7.3c shows a simulated STEM image of the reconstructed Si surface, and line profiles across the step (Fig. 7.3d) show that the simulation reproduces the observed intensity variations well, and that there is certainly a change in relative intensity after 55 s of irradiation under the beam.

#### 7.4.2 Sample preparation and the “clean” Si(110) surface

It is important to recognise that the sample fabrication in this work differs from the usual approach taken to prepare a clean, atomically-flat surface for STM studies [313, 314]. Silicon in particular is a difficult specimen to fabricate due to problems with contamination by elements such as oxygen and carbon—both of which can be found in the STEM column even under high vacuum. Typical STM experimental conditions require vacuum pressure several orders of magnitude less than even the best STEM instruments [315], and great care needs to be taken at all stages of fabrication to avoid contamination. A common method for preparing a clean Si surface at low temperatures for STM experiments is to use a wet chemical process [315], etching with either HF or  $\text{NH}_4\text{F}$  to produce a H-passivated surface. Alternative approaches involve heating the sample to high temperatures (1000 K or more) at very low pressures [313].

Sputtering using 300 eV Ar ions and a subsequent anneal at 1000 K was performed in Ref. 316 to prepare a clean and flat Si(001) surface, and it is worth noting that the method described in Section 7.3 involved 900 eV Ar ion sputtering. However, there was no subsequent annealing, and the problem of achieving truly ultrahigh vacuum conditions remains. Without further, controlled experiments (possibly combining STEM and STM studies), there is no way to confidently identify the cause of the features seen in the STEM sequences. Perhaps the most likely explanation is that we are observing surface oxide structures, linked to the underlying crystalline Si in  $\langle 112 \rangle$  directions as per Ref. 306, that are heavily influenced by the electron beam and so develop over time. The following section investigates the initial stages of oxidation on the Si(110) surfaces.

#### 7.4.3 Oxidation of the Si(110) surface

Atomistic modelling can be used to explore whether the distinct  $\langle 112 \rangle$  features observed are caused by the Si surface itself, or by the formation or dynamics of an oxide layer. MD simulations should also help explain the evolution of the features over time. However, a fundamental problem in MD simulations is accurately determining the

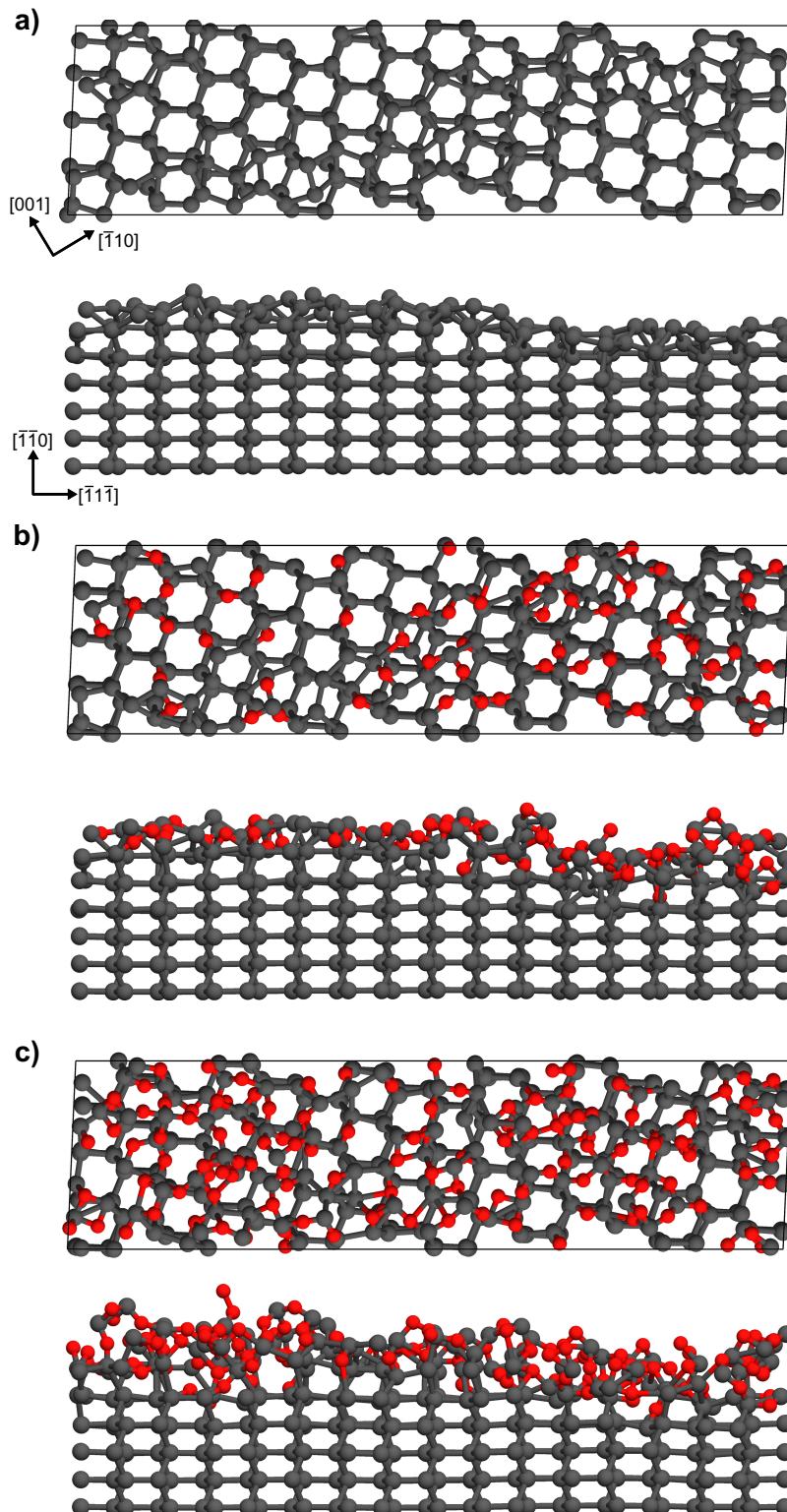
forces on each atom in order to evolve the system in time. Using a first-principles method such as DFT to calculate the forces is the preferred option for surfaces such as Si(110), since dangling bonds and van der Waals interactions with adsorbates can have a significant effect on the stability of the structure. However, while DFT has been used to successfully investigate the energy of the  $16\times 2$  reconstruction [302, 303], the size of the unit cell severely limits the feasibility of *ab initio* MD simulations due to the computational cost of repeated force calculations.

Classical MD simulations offer a much faster method for investigating the dynamics of large surfaces, since empirical potentials and forces can be evaluated much faster than the first principles treatment of DFT. Classical potentials can range from a simple Lennard-Jones model to more complex methods incorporating directionality and long-range interactions [317]. Selecting the appropriate potential is extremely important for obtaining an accurate end result, and this usually means using a potential that has been fitted to experimental properties of the system in question. For instance, a potential that reproduces the bulk properties of Si but not the surface properties is of little use for the present study. Fortunately, the importance of Si in the semiconductor industry means that a number of potentials exist for accurately modelling both bulk and surface properties. These range from the well-established Stillinger-Weber (SW) [318] and Tersoff [319] potentials to more recent approaches such as the Modified Embedded-Atom Method (MEAM) [320, 321]. Each has its own benefits and drawbacks; the SW and Tersoff models are extremely quick to evaluate, while MEAM better reproduces properties such as thermal expansion, as well as the structure of amorphous Si.

A key requirement in this study is that the potential can model the initial oxidation of the Si surface. This is more complicated than the pure Si case, since various inter-atomic interactions, including non-bonding interactions such as van der Waals forces, need to be accounted for. The reactive force field (ReaxFF) approach of van Duin et al. is designed with these components in mind, and is aimed at bridging the gap between expensive, small-scale DFT calculations and faster, large-scale empirical MD simulations [322]. The present work uses the ReaxFF potential developed specifically for the Si–O system [323], which successfully describes the DFT-calculated surface reconstruction shown in Figure 7.1. The ReaxFF potential also replicates the spontaneous dissociation of O<sub>2</sub> into atomic oxygen on the surface of silicon, which was found using *ab initio* MD to be the first step in the oxidation process on other low-index Si surfaces [324].

The simulations were performed using the Large-scale Atomic/Molecular Massively Parallelized Simulator (LAMMPS) code [325, 326]. Starting from the DFT-optimized structure [303], a  $16\times 2$  slab with 14 atomic layers and 20 Å of vacuum was constructed. The bottom layer of Si atoms were fixed to the bulk lattice parameters as calculated with ReaxFF, and the next 7 layers were rescaled to a constant temperature to provide





**Figure 7.4:** Effect of oxidation on the Si(110) surface. **(a)** The clean  $16 \times 2$  surface reconstruction (cf. Fig. 7.1). **(b)** Si(110) surface after the adsorption of 64 O atoms (1 ML coverage) simulated over a period of 250 ps at 1000 K. **(c)** Si(110) surface after the adsorption of 128 O atoms (2 ML coverage). Si atoms are grey, O atoms are red.

a thermal bath for the system. The remaining layers were left unconstrained. An initial energy minimization was carried out using the conjugate-gradient method. Following the procedure of Pamungkas et al. [327], the clean surface was equilibrated at 1000 K for 10 ps using a Langevin thermostat and a timestep of 0.25 fs. The high temperature mimics the effect of local sample heating by the electron beam. An O<sub>2</sub> molecule was then introduced every 5 ps at a random point above the surface with a small downward velocity to model the oxidation process, up to a total of 64 O atoms, corresponding to approximately 1 monolayer (ML) coverage of the surface. The same oxidation procedure was repeated with 128 O atoms (i.e. 2 ML coverage). A final energy minimization was performed to fully relax the system in all cases. For the ADF-STEM simulations, the unit cells from the MD simulations were extended in the direction of the bulk crystal to form a slab 7 nm thick, thus matching the estimated lamella thickness.

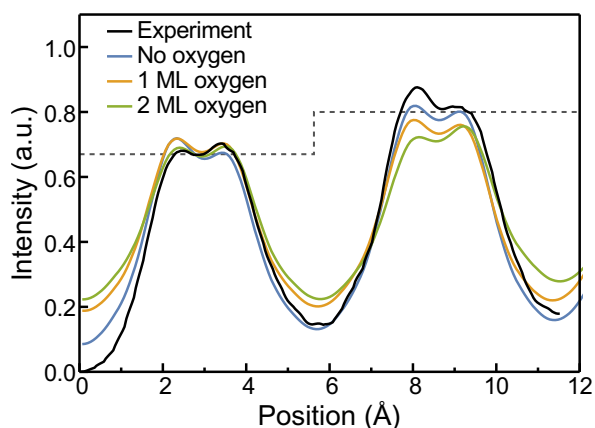
The fully-relaxed surfaces shown in Figure 7.4 demonstrate that as the oxygen coverage increases, the step seen in the  $\langle 112 \rangle$  direction is gradually lost. In Figure 7.4c oxygen atoms are starting to diffuse into the bulk Si, and ultimately a layer of SiO<sub>2</sub> will form as more oxygen is deposited, thus removing the  $\langle 112 \rangle$  feature entirely and leaving behind an Si/SiO<sub>2</sub> interface [308].

Figure 7.5 shows line profiles taken across the  $\langle 112 \rangle$  step in each of the simulated ADF-STEM images (see highlighted region in Fig. 7.3c). The Si dumbbells are resolved in all three cases, and the profiles for 1 and 2 ML coverage exhibit very similar behaviour. The presence of oxygen is seen to reduce the change in intensity across the step, and also leads to an increase in intensity between the dumbbells as a result of the disordered surface structure (cf. Fig. 7.4b,c). This suggests that the bright chevron-like features in the  $\langle 112 \rangle$  direction (Fig. 7.3b) become more visible as oxygen is sputtered from the surface under the electron beam, and adds weight to the hypothesis that we are observing the dynamics of surface oxide structures linked to the underlying crystalline Si in the  $\langle 112 \rangle$  directions.

#### 7.4.4 Separating the top and bottom surfaces

Since STEM is a projected imaging method, one must remember that the observed features may exist on both the top and bottom surfaces, and that separating their contributions to the image intensity can be difficult. However, Figure 7.5 showed that the oxidized Si surface does not exhibit the same bright chevron-like features as cleaner surfaces, indicating that the interaction between the surface and the electron beam leads to the surface changing, and thus the contrast becoming clearer over time in the experimental sequences. Previous work has shown that oxygen is preferentially sputtered from SiO<sub>2</sub> by the electron beam [312], and any sputtering is also likely to





**Figure 7.5:** Line profiles across the surface step in the simulated  $16\times 2$  reconstruction of Si(110) are compared with the experimental line profile (after 55 s of imaging, see Fig. 7.3d) for a clean Si surface, 1 ML coverage of oxygen and 2 ML coverage. The dashed grey line indicates the position of the step on the clean surface.

occur first on the bottom surface of the lamella, with the oxygen coverage on the top surface expected to be more stable.

Further quantification of the oxygen sputtering was conducted using electron energy-loss spectroscopy (EELS). EELS spectra were acquired using an FEI Titan<sup>3</sup> operating at 300 kV and equipped with a GIF Quantum ERS system. Spectra were recorded every 2 s over an area of  $1500\text{ nm}^2$ , and the probe current was 50 pA. The O K edge intensity was calculated using Gatan Digital Micrograph. The background was removed by pre-edge power-law fitting, and quantification was done by using a Hartree-Slater cross-section model for each spectrum of the time series. Figure 7.6 demonstrates that the oxygen content of the sample as measured with EELS decreases over time due to exposure to the electron beam. After 60 seconds the oxygen content is approximately half the initial value. Although not conclusive, this analysis points towards oxygen being sputtered off one surface of the lamella by the beam, leaving behind an altered Si(110) surface, as seen in the ADF-STEM image sequences. The oxygen content on the top surface is suggested to remain relatively stable relative to the bottom, but further work would be needed to confirm this hypothesis.

As a final note, acquiring a tilt series or a through-focal series are alternative STEM methods to provide depth resolution and confirm the contributions of the top and bottom surfaces, but both come at the expense of temporal resolution and possibly a higher cumulative electron dose. Furthermore, the depth resolution of a through-focal series may not be sufficient to distinguish the ca.  $1\text{ Å}$  step height on the Si(110) surface.

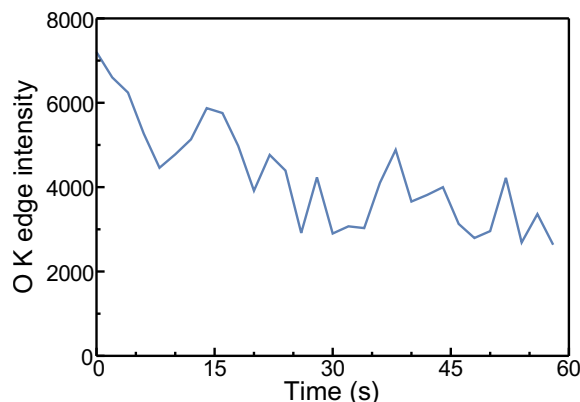


Figure 7.6: Oxygen K edge intensity from an EELS time series.

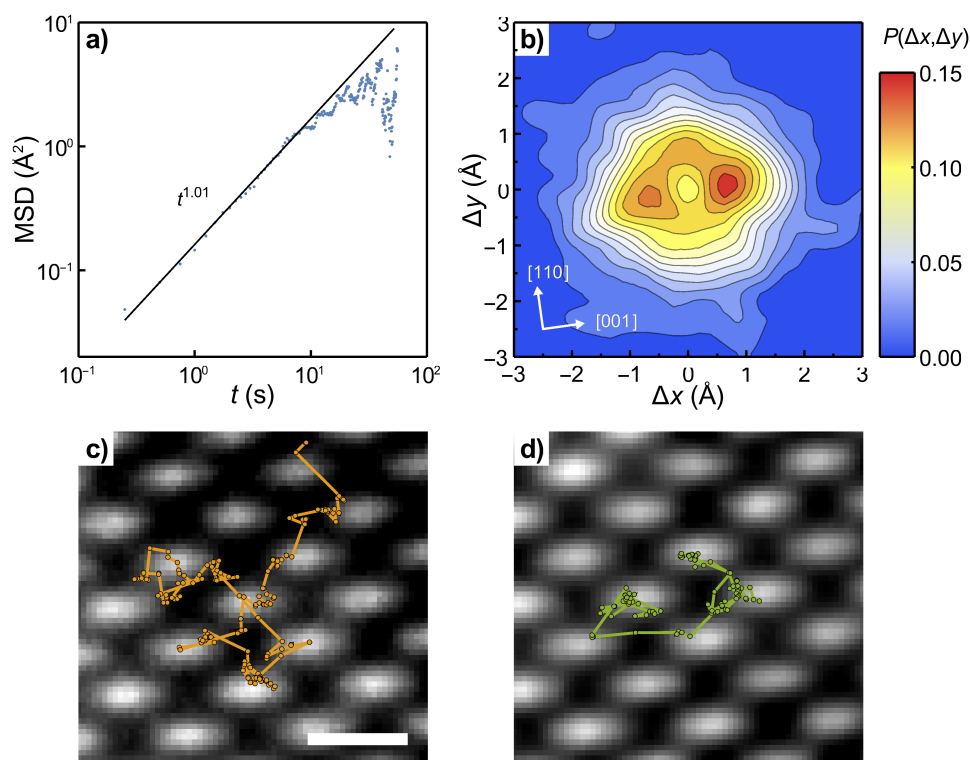
## 7.5 Pt adatom diffusion on Si(110)

We now turn our attention to the sparse component of the RPCA decomposition (Fig. 7.2d) and investigate the dynamic behaviour of the Pt adatoms on the Si surface. At the experimental accelerating voltage of 300 kV, a collision with an electron can impart up to 4.5 eV to a Pt atom (Eq. 2.2), which will be more than sufficient to excite the adsorbed atom out of a surface site and thus diffuse over the surface.

The wavelet-based feature detector developed in Chapter 4 was used to extract the approximate positions of the adatoms from one of the sequences, and the positions were further refined by fitting 2D Gaussians to a small window around each atom. The use of the classification-based feature detector is necessary here to account for the imperfect background subtraction, motion blur and noise. The positions were then combined into trajectories using the reinforcement learning approach again developed in Chapter 4; in total, 309 particle tracks were analyzed with lengths of 2–56 s.

The MSD of the trajectories is seen to be linear in time (Fig. 7.7a), with an exponent of  $1.01 \pm 0.02$ , indicating that the Pt adatoms are undergoing a normal diffusion process as a result of the beam-induced motion. This is in stark contrast to the observation of anomalous subdiffusion in Chapter 5, and reflects the relative homogeneity of the Si substrate compared to the disordered structure of graphene oxide. Clearly the possible presence of oxygen and other adsorbates is insufficient to noticeably alter the diffusion behaviour of Pt. Note that at long times the noise in the MSD increases due to a smaller number of trajectories available for averaging. Applying the tests for anomalous diffusion from Chapter 5, no ageing is observed in the system, and the results point towards a simple “random walk on a lattice” model.

Analysis of the magnitude and direction of the atom displacements in Figure 7.7b reveals a preference for left-to-right adatom jumps in approximately the [001] direction. Figure 7.7b plots the probability of a displacement,  $P(\Delta x, \Delta y)$ , weighted by the



**Figure 7.7:** (a) MSD of the atom trajectories extracted from the ADF-STEM sequence. The best-fit line is shown in black along with the fitted exponent. (b) The probability of adatom displacements,  $P(\Delta x, \Delta y)$ , weighted by the displacement magnitude. (c,d) Two of the particle tracks overlaid onto the background image of the Si lattice (Fig. 7.2c). Scale bar: 0.5 nm.

magnitude to give additional importance to long jumps (which are likely beam-driven) and reduce the influence of very small displacements (which are most likely caused by measurement error).

Two important points can be drawn from this plot. The first is that there is a strong preference for jumps in line with the Si dumbbells, which are tilted a few degrees off the horizontal axis. Figures 7.7c and d overlay two of the atomic trajectories on the substrate image extracted with RPCA (Fig. 7.1c), and show that there is indeed a tendency for the adatom to sit atop a dumbbell rather than off it, although the image resolution is insufficient to elucidate the complete local structure. On the reconstructed surface one Si atom of the pair will sit higher than the other, potentially providing a stable site for the Pt atom. The role of oxygen atoms in binding the Pt atoms to the substrate cannot be ruled out either, and although DFT calculations could resolve this ambiguity over stable sites, the size and complexity of the (oxidized) surface reconstruction and the beam-induced changes make this a potentially challenging computation.

The second observation from Figure 7.7b is the preference for jumps in the [001] direction. This is in the scan direction of the STEM probe, suggesting that the adatoms

are, in effect, being “pushed” by the electron beam. The aberration-corrected probe has a Lorentzian profile and an estimated full-width half-maximum of 1.2 Å. In comparison, the experimental pixel size is 0.24 Å, meaning that a significant fraction of the electron dose will fall outside the pixel being imaged. The probe tails thus increase the likelihood of an adatom being displaced when the probe is nearby. Since the probe will always approach the Pt adatom from the left in this experiment, it is suggested that this contributes to the directional bias, and therefore the scan direction is another important consideration for in-situ studies of adatom dynamics, alongside aspects such as accelerating voltage, beam current and dwell time. Understanding the relationships between beam effects and the underlying physical process is crucial for a full characterization of surface dynamics, and future experiments that vary the scan direction, for example between individual frames as in the RevSTEM technique [328], may shed further light on the origins of the directional bias.

## 7.6 Summary

This chapter has demonstrated a novel approach to observing and understanding the dynamics of surfaces using atomically-resolved STEM image sequences, combining the image processing methods outlined in Chapters 3 and 4. By exploiting temporal correlations between frames using the ORPCA algorithm, it is possible to recover the dynamic surface behaviour of Si(110), and relate the observations to computational modelling of surface oxide layers. The separation of the low-rank and sparse components of the time series facilitated further analysis of the dynamics of Pt surface atoms, with motion induced by the electron beam revealing a directional bias attributable to the STEM probe scan direction. The analysis presented here extends the methods in this thesis to crystalline substrates, and shows that they are widely applicable to the study of atomic-scale surface dynamics with aberration-corrected STEM. Crucially, combining quantitative ADF imaging with MD simulations maintains both the temporal resolution and the low dose imaging conditions, which are key criteria for successfully observing single-atom dynamics.

## Chapter 8

# Conclusions and future work

### 8.1 Summary

This thesis has explored the potential of STEM as a technique for time-resolved materials characterization with atomic resolution. Through a series of case studies, it has shown how image processing, machine learning and computational modelling can be combined with aberration-corrected electron microscopy to explore the dynamic behaviour of atoms on surfaces, and, crucially, to begin to understand the effects of the beam-specimen interaction at the single-atom level.

Chapter 3 addressed some of the challenges faced in handling and analyzing the large and noisy EM datasets by exploiting correlations using novel low-rank matrix techniques. Focusing on time-resolved imaging, a new algorithm for denoising image sequences under challenging noise conditions was developed. Consideration of the statistics of the signal acquisition process was highlighted as an important factor in implementing an objective, automated method for parameter selection. The investigation of low-rank methods was extended using an online robust PCA algorithm, showing the feasibility of real-time denoising, dimensionality reduction and background subtraction for immediate feedback on electron microscopy experiments.

The benefits of incorporating prior knowledge into analytical tools were discussed in Chapter 4, which presented adaptable computational methods for detecting and tracking the motion of atoms in image sequences. The signal acquisition process was again analyzed to understand the detection limits for single atom imaging in the presence of noise, which is an important factor in low-dose STEM imaging. Together these techniques can guide experimental design to maximize the useful information content acquired whilst guarding against significant damage by the electron beam.

Chapter 5 presented an initial study of the potential of time-resolved STEM imaging, showing how Cu atoms adsorbed onto GO undergo anomalous diffusion as a result of

the spatial and energetic heterogeneity present in the substrate. Extensive simulations revealed that GO can be modelled using percolation theory, resulting in a fractal-like surface that was explored further in Chapter 6. The insights gained about the binding of single Cu atoms to GO were extended to the study of small Cu clusters on GO, where *ab initio* random structure searching revealed the importance of accurately modelling the substrate when considering the structure of small metal clusters.

Finally, in Chapter 7, the image analysis methods developed previously were applied to the study of a crystalline system. This illustrated the power of quantitative ADF imaging and image simulation by revealing the evolution of surface structures on Si(110) over time. Complementary analysis of the motion of Pt adatoms on the surface was achieved using background subtraction with robust PCA, and showed how the electron beam can influence the behaviour of adatoms.

These case studies demonstrate the novel materials insights that can be elucidated using time-resolved STEM, with applications such as the study of heterogeneous catalysts, surface reconstruction and growth, and the fundamentals of atomic diffusion. Crucially, this thesis establishes a robust workflow for the qualitative and quantitative characterization of materials dynamics with electron microscopy. The combination of state-of-the-art imaging with computational analysis and atomistic modelling provides an objective, automated framework for exploring important physical and chemical behaviours at the atomic scale.

## 8.2 Future work

Two broad aspects of the work presented in this thesis offer potential for further research. These are the development of methods for real-time data analysis, and the application of time-resolved (S)TEM techniques to in-situ experiments to investigate the dynamic behaviours of single atoms under reaction conditions.

### 8.2.1 Real-time data analysis in electron microscopy

Chapters 3 and 4 introduced a series of techniques for processing and analyzing time-resolved (S)TEM data. A natural extension would be to combine the denoising, feature extraction and atom tracking steps into a single, seamless routine that can be integrated directly with the computer controlling the microscope. This would provide real-time feedback on experiments and make adjusting the parameters to suit the sample, such as the electron dose or focus, very straightforward. Successfully constructing such a system is ultimately a software development challenge, and would require collaboration with manufacturers to interface with both the microscope hardware and the typically

closed-source control software. This could include the integration of more powerful computing resources, perhaps using GPUs or through links to cloud-based services akin to the distributed computing model in particle physics [329]. Briefly moving away from single-atom imaging, it was noted in Chapter 3 that robust, online unsupervised learning is not limited to the ORPCA algorithm. Exploring the potential for online NMF for processing EELS and EDS datasets in real-time is another avenue worth pursuing [167], and demonstrating the general applicability of these techniques may further encourage collaboration.

Returning to single atoms, extending the particle detection and tracking algorithms of Chapter 4 to work with multiple elements would facilitate the investigation of the structure and dynamics of bimetallic clusters, thus building on the work with copper on GO presented in Chapter 6. Use of ABF imaging in conjunction with ADF-STEM could enable the simultaneous, real-time tracking of heavier metal atoms alongside lighter atoms such as oxygen during catalytic reactions. Plotting diffusion behaviours such as the MSD as the data is acquired would highlight any significant changes in adatom dynamics, for example the point at which the accumulated radiation damage leads to a change in the diffusion exponent.

Lastly, an interesting point worth investigating is whether it is possible to recover the 3D structure of small atomic clusters with temporal resolution. A recent study of sub-nanometre Ge clusters used ADF-STEM and statistical atom-counting to construct an initial model of the cluster in consecutive frames of a sequence. Each model was subsequently relaxed with DFT [330]. This final step is clearly not “real-time” by any means, and Ref. 330 largely ignores the cluster–substrate interaction in the calculations, in contrast to the findings in Chapters 5 and 6. Pursuing alternatives is therefore of significant interest. One possible solution may come from computer vision methods more commonly found in self-driving cars, where the challenge is to build a 3D model of the environment from a sequence of 2D images by identifying common (or corresponding) points and features in each frame. This is known as “structure from motion” (SfM) [331]. It is possible to reconstruct simple shapes even when the feature correspondence between frames is unknown using an expectation-maximization algorithm [332]. This is particularly relevant to atomic clusters since distinguishing between atoms with the same mass is impossible using Z-contrast alone. Imposing sensible constraints on interatomic spacing, as well as prior knowledge obtained from DFT calculations, could make the SfM technique a powerful tool for analyzing the dynamics of small atomic clusters in three dimensions, building on work on the reversible dynamics of a Si<sub>6</sub> cluster in a graphene pore [109].

### 8.2.2 In-situ experiments with single atom catalysts

The second major theme for future work focuses on the materials science questions outlined in Chapters 5–7. The work in these chapters represents a series of case studies demonstrating the potential applications of the image processing and data analysis techniques of the first part of the thesis. The future directions are numerous, but focusing on the dynamics of single atoms, a key consideration not yet explored is the effect of temperature and the interplay between thermal and beam-driven effects. Controlling and analyzing these effects using STEM heating holders is an obvious next step in studying single atom dynamics with electron microscopy. Going even further, environmental STEM raises the prospect of following catalytic reactions in the presence of gases with unprecedented resolution [33–36].

Of course, achieving these goals will require a systematic study of the effects of the beam dose and accelerating voltage to determine the point at which the beam interaction is minimized, while the various image processing methods are still able to recover the motion of single atoms. Chapter 7 suggested that the beam scan direction may play an influential role, and conducting further studies involving different raster scan patterns would clarify this effect.

Having addressed the optimal conditions for single atom imaging with in-situ microscopy, it would be very interesting to investigate the anomalous diffusion behaviour first discussed in Chapter 5. In that case subdiffusion was observed, raising the question of whether superdiffusion at the atomic scale is also possible, and what the necessary characteristics of the system are for it to arise. In a recent paper, Barry et al. showed that Os atoms deposited on boronic graphene doped with Se moved faster under the electron beam than on the same material doped with S [333], suggesting that controlling atomic motion at surfaces through doping is a promising idea for future study. Although no MSD analysis was presented in Ref. 333, the jump distributions were shown to be inconsistent with Brownian motion, with more long-range jumps than expected under normal diffusion.

Returning to subdiffusion, the anomalous motion of Cu on GO was caused by a CTRW model combined with a fractal-like substrate, resulting in both spatial and energetic disorder. Investigating other systems with similar disorder would be very important for heterogeneous catalysis, especially with the recent trend towards developing single atom catalysts [23]. Understanding the adatom–substrate interaction is crucial to controlling and tuning the properties of a catalyst, and preventing the aggregation of atoms is key to preventing the degradation of the system. Exploring the properties of other technologically-important substrates such as graphitic carbon nitride may prove enlightening [31, 262], since the fractal properties of GO discussed here may place a limit on the tunability of the substrate.



---

Finally, one of the ultimate goals of aberration-corrected STEM would be to control materials at the single atom level, in much the same way as STM can be used to carefully place atoms into small structures on surfaces [334–336]. Of course, understanding and carefully controlling the beam–specimen interaction is the major outstanding issue. Combining the real-time analysis techniques in this thesis with systematic studies of the electron beam effects will constitute a significant step towards overcoming this barrier.



# Appendix A

## Supplementary Materials

### A.1 Supplementary Movies

The following movies are available on the supplementary CD. Movies S1–7 and S13 are reproduced from Ref. 117.

**Movie S1–6** The *Microtubule*, *Vesicle*, *Receptor*, *Microtubule+clouds*, *Vesicle+clouds* and *Receptor+clouds* test sequences.

**From left:** Original images; corrupted with Poisson–Gaussian noise; denoised with V-BM4D; denoised with PGURE-SVT.

**Movie S7** ADF-STEM sequence of a nanoparticle on graphene oxide, before (left) and after (right) denoising with PGURE-SVT. Sequence recorded and played back at 4 fps. The field of view is 6.8 nm.

**Movie S8** Comparison between the block coordinate descent (BCD) and stochastic gradient descent (SGD) methods of the ORPCA algorithm for denoising STEM image sequences including sample drift.

**From left:** Clean images; corrupted with Poisson–Gaussian noise; denoised using the BCD method; denoised using the SGD method.

**Top row:** Simulated ADF sequence of  $\text{SrTiO}_3$ .

**Bottom row:** Simulated ABF sequence of  $\text{SrTiO}_3$ .

**Movie S9** Application of ORPCA for the inpainting of undersampled STEM image sequences including sample drift.

**From left:** Clean images; corrupted with Poisson–Gaussian noise and with only 25% of pixels observed; after processing with ORPCA.

**Top row:** Simulated ADF sequence of  $\text{SrTiO}_3$ .

**Bottom row:** Simulated ABF sequence of  $\text{SrTiO}_3$ .

**Movie S10** Foreground–background separation using the ORPCA algorithm, in the presence of sample drift.

**Top row:** Simulated ADF sequence of a Si dopant atom in graphene, and after corruption with Poisson–Gaussian noise.

**Middle row:** Background and foreground images recovered with the ORPCA algorithm.

**Bottom row:** Mean image from the noisy sequence after alignment, and the difference image between the mean and each frame of the aligned sequence.

**Movie S11** ADF-STEM sequence 1 of copper adatoms on graphene oxide, before and after denoising with PGURE-SVT. Sequence recorded and played back at 8 fps. The field of view is 6.8 nm.

**Movie S12** ADF-STEM sequence 2 of copper adatoms on graphene oxide, before and after denoising with PGURE-SVT. Sequence recorded and played back at 8 fps. The field of view is 6.8 nm.

**Movie S13** ADF-STEM sequence of Cu adatoms on graphene oxide adopting a trimer motif, before and after denoising with PGURE-SVT. Sequence recorded and played back at 8 fps. The field of view is 4.8 nm.

**Movie S14** Sequence 1 of Pt adatoms on Si(110) imaged using ADF-STEM. Originally recorded at 4 fps, playback has been sped up 4×. The field of view is 5.4 nm.

**Top row:** As-acquired images, and after denoising and alignment.

**Bottom row:** Background and foreground after processing with ORPCA.

**Movie S15** Sequence 2 of Pt adatoms on Si(110) imaged using ADF-STEM. Originally recorded at 4 fps, playback has been sped up 4×. The field of view is 4.2 nm.

**Top row:** As-acquired images, and after denoising and alignment.

**Bottom row:** Background and foreground after processing with ORPCA.

## A.2 Supplementary Software

The following programs were developed as part of this thesis, and have been made freely available under an open-source license. The source code for PGURE-SVT and CTRWfractal is also provided on the supplementary CD:

**PGURE-SVT** was written in C++ using the Armadillo linear algebra library [337], and includes a wrapper for use with the HyperSpy software package [338]. The code is available to download from <http://tjof2.github.io/pgure-svt>.

**ORPCA** was written in Python, based on code provided by the original authors of the algorithm [163], and includes the initialization and gradient descent adaptations described in Section 3.4. The code has been contributed to the HyperSpy software package [338].

**CTRWfractal** was written in C++ and adapted to simulate continuous-time random walks from work on percolation by Newman & Ziff [339]. The code is available to download from <http://github.com/tjof2/ctrwfractal>.



## Appendix B

# Poisson–Gaussian unbiased risk estimator

The derivation of the Poisson–Gaussian estimator assumes a mixed noise model described in Equation 3.12. For notational convenience, here matrices are replaced with vectors, which is equivalent to flattening the images. The noise model is then:

$$\mathbf{y} = \alpha \mathbf{z} + \mathbf{e} \text{ with } \begin{cases} \mathbf{z} \sim \mathcal{P}(\frac{\mathbf{x}}{\alpha}) \\ \mathbf{e} \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \quad (\text{B.1})$$

This model can be rewritten as  $\mathbf{y} = \alpha \mathbf{z} + \mathbf{b} + \mu$ , where  $\mathbf{b} \sim \mathcal{N}(0, \sigma^2)$ . The derivation in this section builds on the work of Le Montagner et al. [144], which set  $\mu = 0$ . The derivation makes use of two known results for Gaussian and Poisson-distributed data, which are restated here.

### B.1 Stein’s Lemma

If  $\mathbf{y} = \mathbf{x} + \mathbf{b}$ , where  $\mathbf{b} \sim \mathcal{N}(0, \sigma^2)$ , and  $\phi$  is a weakly differentiable function such that  $\mathbb{E} \left\{ \sum_n \left| \frac{\partial \phi_n}{\partial y_n}(\mathbf{y}) \right| \right\} < +\infty$  for all  $n$ , then from Ref. 136:

$$\mathbb{E} \{ \langle \mathbf{b} | \phi(\mathbf{y}) \rangle \} = \sigma^2 \mathbb{E} \{ \text{Div } \phi(\mathbf{y}) \} \quad (\text{B.2})$$

where  $\text{Div } \phi(\mathbf{y}) = \sum_n \frac{\partial \phi_n}{\partial y_n}(\mathbf{y})$ .

## B.2 Hwang’s Lemma

In Ref. 340, Hwang derived an equivalent of Stein’s Lemma for Poisson-distributed data. For a function  $\psi(x)$ , where  $\mathbb{E}\{|\psi(x)|\} < +\infty$ , and  $x \sim \mathcal{P}(\lambda)$ :

$$\mathbb{E}\{\lambda\psi(x)\} = \mathbb{E}\{x\psi(x-1)\} \quad (\text{B.3})$$

This result can be extended to vectors by setting  $\mathbf{z} \sim \mathcal{P}(\mathbf{x})$ , and  $\psi$  be a function such that  $\mathbb{E}\{\sum_n |\psi_n(\mathbf{z})|\} < +\infty$  for all  $n$ :

$$\mathbb{E}\{\langle \mathbf{x} | \psi(\mathbf{z}) \rangle\} = \mathbb{E}\left\{\left\langle \mathbf{z} | \psi^{[-1]}(\mathbf{z}) \right\rangle\right\} \quad (\text{B.4})$$

where  $\psi^{[-1]}(\mathbf{z}) = [\psi_n(\mathbf{z} - \mathbf{c}_n)]_{n=1\dots N}$ , and  $\mathbf{c}_n$  is a vector of zeros except for the  $n^{\text{th}}$  element, which is set to 1.

## B.3 Poisson–Gaussian unbiased risk estimator

The results in Equation B.2 and Equation B.4 can now be used to rewrite the expectation of the mean-squared error (Eq. 3.9) in terms of the observation vector  $\mathbf{y}$  only:

$$\begin{aligned} \mathbb{E}\{\text{MSE}\} &= \frac{1}{N} \mathbb{E}\left\{\|f(\mathbf{y}) - \mathbf{x}\|_2^2\right\} \\ &= \frac{1}{N} \left( \mathbb{E}\left\{\|f(\mathbf{y})\|_2^2\right\} - 2\mathbb{E}\{\langle \mathbf{x} | f(\mathbf{y}) \rangle\} + \|\mathbf{x}\|_2^2 \right) \\ &= \frac{1}{N} \left( \mathbb{E}\left\{\|f(\mathbf{y})\|_2^2\right\} - \underbrace{2\mathbb{E}\{\langle \mathbf{x} | f(\mathbf{y}) \rangle\}}_{\text{I}} + \underbrace{\mathbb{E}\{\langle \mathbf{x} | \mathbf{y} \rangle\}}_{\text{II}} \right) \end{aligned} \quad (\text{B.5})$$

Applying Equation B.4 to term I (cf. Ref 144):

$$\begin{aligned} \mathbb{E}\{\langle \mathbf{x} | f(\mathbf{y}) \rangle\} &= \mathbb{E}\left\{\alpha \mathbb{E}_{\mathbf{z}}\left\{\left\langle \frac{\mathbf{x}}{\alpha} | f(\alpha \mathbf{z} + \mathbf{b} + \mu) \right\rangle\right\}\right\} \\ &= \mathbb{E}\left\{\alpha \mathbb{E}_{\mathbf{z}}\left\{\left\langle \mathbf{z} | f^{[-1]}(\alpha \mathbf{z} + \mathbf{b} + \mu) \right\rangle\right\}\right\} \\ &= \mathbb{E}\left\{\left\langle \mathbf{y} - \mathbf{b} - \mu | f^{[-\alpha]}(\alpha \mathbf{z} + \mathbf{b} + \mu) \right\rangle\right\} \\ &= \mathbb{E}\left\{\left\langle \mathbf{y} | f^{[-\alpha]}(\mathbf{y}) \right\rangle\right\} - \mathbb{E}\left\{\left\langle \mathbf{b} | f^{[-\alpha]}(\mathbf{y}) \right\rangle\right\} - \mathbb{E}\left\{\left\langle \mu | f^{[-\alpha]}(\mathbf{y}) \right\rangle\right\} \end{aligned} \quad (\text{B.6})$$

Then applying Equation B.2:

$$\mathbb{E}\{\langle \mathbf{x} | f(\mathbf{y}) \rangle\} = \mathbb{E}\left\{\left\langle \mathbf{y} | f^{[-\alpha]}(\mathbf{y}) \right\rangle - \sigma^2 \text{Div } f^{[-\alpha]}(\mathbf{y}) - \mu f^{[-\alpha]}(\mathbf{y})\right\} \quad (\text{B.7})$$



For term II, we replace  $f(\mathbf{y})$  in Equation B.7 by the identity function:

$$\mathbb{E} \{ \langle \mathbf{x} | \mathbf{y} \rangle \} = \mathbb{E} \{ \|\mathbf{y}\|_2^2 - (\alpha + \mu) \mathbf{y} \} + \mu - N\sigma^2 \quad (\text{B.8})$$

Putting everything together according to Equation B.5:

$$\begin{aligned} \text{PGURE} = \frac{1}{N} \left( \|f(\mathbf{y})\|_2^2 + \|\mathbf{y}\|_2^2 - 2 \langle \mathbf{y} | f^{[-\alpha]}(\mathbf{y}) \rangle + 2\sigma^2 \text{Div } f^{[-\alpha]}(\mathbf{y}) \right. \\ \left. + 2\mu f^{[-\alpha]}(\mathbf{y}) - (\alpha + \mu) \mathbf{y} + \mu \right) - \sigma^2 \quad (\text{B.9}) \end{aligned}$$

The next step is to apply a first-order Taylor expansion:

$$f^{[-\alpha]}(\mathbf{y}) \approx f(\mathbf{y}) - \alpha \partial f(\mathbf{y}) \quad (\text{B.10})$$

This simplifies Equation B.9:

$$\begin{aligned} \text{PGURE} = \frac{1}{N} \left( \|f(\mathbf{y}) - \mathbf{y}\|_2^2 + 2\mu f(\mathbf{y}) + 2 \langle (\alpha(\mathbf{y} - \mu) + \sigma^2) | \partial f(\mathbf{y}) \rangle \right. \\ \left. - 2\alpha\sigma^2 \partial^2 f(\mathbf{y}) - (\alpha + \mu) \mathbf{y} + \mu \right) - \sigma^2 \quad (\text{B.11}) \end{aligned}$$

Replacing  $\mathbf{y}$  by  $\mathbf{Y}$ , and the element-wise product  $\langle \mathbf{a} | \mathbf{b} \rangle$  by its matrix equivalent  $\mathbf{A} \circ \mathbf{B}$ , we arrive at Equation 3.13 with  $\mathbb{E} \{ \text{MSE} \} = \mathbb{E} \{ \text{PGURE} \}$ .



## Appendix C

# Stochastic gradient descent

### C.1 Theory

Gradient descent methods can be used to improve the convergence of the L-update step in the ORPCA algorithm. First recall that Equation 3.21 has the form:

$$f(x) = \sum_{i=1}^t f_i(x) \quad (\text{C.1})$$

Determining the local minimum of this function with a standard batch gradient descent method involves taking the negative of the gradient at the current point:

$$x := x - \gamma \sum_i^t \nabla f_i(x) \quad (\text{C.2})$$

where  $\gamma$  is the step size. As seen with the closed-form and block coordinate descent methods presented in Chapter 3, this approach requires a running total of the previous samples and so will struggle to adapt quickly to any changes in the function over time. In contrast, stochastic gradient descent seeks the minimum by estimating the gradient from just a single sample [341]:

$$x := x - \gamma \nabla f_i(x) \quad (\text{C.3})$$

Stochastic gradient descent provides a noisy approximation to the true gradient of the function, but with improved computational performance and a faster response to changes in the function. A trade-off can be made to reduce noise by using small batches of samples in Equation C.2 rather than all previous samples.

## C.2 Derivation of the update step

Applying either gradient descent method requires the derivative:

$$\frac{\partial}{\partial \mathbf{L}} \left( \frac{1}{2} \|\mathbf{y}_i - \mathbf{L}\mathbf{r}_i - \mathbf{s}_i\|_2^2 + \lambda \|\mathbf{r}\|_2^2 + \xi \|\mathbf{s}\|_1 + \frac{\lambda}{2} \|\mathbf{L}\|_F^2 \right) \quad (\text{C.4})$$

To calculate Equation C.4, we make use of the following identities. First, the  $l_2$ -norm of a vector  $\mathbf{x}$  and the Frobenius norm of a matrix  $\mathbf{X}$  can be rewritten as:

$$\|\mathbf{x}\|_2^2 = \mathbf{x}^T \mathbf{x} = \text{Tr} [\mathbf{x}\mathbf{x}^T] \quad (\text{C.5})$$

$$\|\mathbf{X}\|_F^2 = \text{Tr} [\mathbf{X}\mathbf{X}^T] \quad (\text{C.6})$$

where  $\text{Tr} [\mathbf{X}]$  is the matrix trace. The second identity is Equation 119 in Ref. 342:

$$\frac{\partial}{\partial \mathbf{X}} \text{Tr} [(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C})(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C})^T] = 2\mathbf{A}^T (\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C}) \mathbf{B}^T \quad (\text{C.7})$$

Defining  $\mathbf{A} = \mathbf{I}$  (the identity matrix), and substituting  $\mathbf{c} = \mathbf{y}_i - \mathbf{s}_i$ , we have:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{L}} \left( \frac{\lambda}{2} \|\mathbf{L}\|_F^2 \right) &= \lambda \mathbf{L} \quad (\text{C.8}) \\ \frac{\partial}{\partial \mathbf{L}} \|\mathbf{y}_i - \mathbf{L}\mathbf{r}_i - \mathbf{s}_i\|_2^2 &= \frac{\partial}{\partial \mathbf{L}} \|\mathbf{c} - \mathbf{L}\mathbf{r}_i\|_2^2 \\ &= \frac{\partial}{\partial \mathbf{L}} \text{Tr} [(\mathbf{c} - \mathbf{L}\mathbf{r}_i)(\mathbf{c} - \mathbf{L}\mathbf{r}_i)^T] \\ &= 2(\mathbf{L}\mathbf{r}_i - \mathbf{c}) \mathbf{r}_i^T \\ &\triangleq 2(\mathbf{L}\mathbf{r}_i - (\mathbf{y}_i - \mathbf{s}_i)) \mathbf{r}_i^T \quad (\text{C.9}) \end{aligned}$$

Putting everything together results in the new update equation for  $\mathbf{L}$ :

$$\mathbf{L} := \mathbf{L} - \gamma (\mathbf{L}\mathbf{r}_i \mathbf{r}_i^T - (\mathbf{y}_i - \mathbf{s}_i) \mathbf{r}_i^T + \lambda \mathbf{L}) \quad (\text{C.10})$$

## C.3 Choosing the step size

Determining the appropriate step size,  $\gamma$ , is an open problem in machine learning, and in practice requires tuning on a small sample of training data\*. Setting  $\gamma$  too small means the algorithm approaches the minima extremely slowly, while if  $\gamma$  is too large the algorithm can oscillate wildly and take a considerable time to settle down. Typical gradient descent algorithms use information about second derivatives of the objective function or expensive line searches to determine the appropriate value of  $\gamma$ . In many high-dimensional cases such as neural network training, the size of the dataset makes

---

\*The step size is also known as the *learning rate*.

the calculation of second derivatives computationally intractable. A common approach is to use a learning rate “schedule”, which decreases  $\gamma$  as each new sample arrives. The general expression used in this work is from Ref. 341:

$$\gamma_t = \gamma_0 (1 + \gamma_0 \lambda t)^{-1} \quad (\text{C.11})$$

which slowly decreases the step size as the number of samples increases, thus reducing the effect each new sample has on the estimate of  $\mathbf{L}$ .

## C.4 Accelerating stochastic gradient descent

Gradient descent methods based on first derivatives alone particularly struggle with saddle points or in long, narrow valleys in the objective function surface. For scenarios where second-derivative information is not available, novel methods have been developed to incorporate memory of the previous steps, such as the momentum method and Nesterov’s accelerated gradient descent [343]. These mitigate the risk of being trapped at saddle points by adding a term based on the direction of the previous update step. For example, the momentum method can be written as:

$$v_t = \eta v_{t-1} + \gamma_t \nabla f_t(x) \quad (\text{C.12})$$

$$x := x - v_t \quad (\text{C.13})$$

where  $\eta$  is a parameter between 0 and 1. Of course, introducing  $\eta$  alongside the step size makes careful parameter selection through training even more important. As discussed in Section 3.4, alternative accelerated methods such as AdaDelta do away with  $\gamma$  altogether [168]. The improved convergence properties and robustness to saddle points of these accelerated methods may prove particularly useful in scenarios where sharp transitions occur between video frames, which are smoothed out by the slower response of standard SGD.



## Appendix D

# Density functional theory

### D.1 Theory

Density functional theory (DFT) is a widely-used approach to calculating the electronic structure of materials [344]. A full quantum mechanical treatment of a system of atoms and electrons involves solving the time-dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} \Psi(\mathbf{r}_j, \mathbf{R}_J, t) = \hat{H} \Psi(\mathbf{r}_j, \mathbf{R}_J, t) \quad (\text{D.1})$$

where  $\mathbf{r}_j$  represents the positions of  $j$  electrons,  $\mathbf{R}_J$  the positions of  $J$  atomic nuclei, and  $\hat{H}$  is the Hamiltonian operator. Solving this problem beyond the simplest of systems is well beyond the scope of modern computational resources. However, since the mass of the nucleus is typically  $10^3$  times larger than the mass of the electron, the wavefunction can be separated into an electronic component  $\psi$  and a nuclear component  $\chi$ :

$$\Psi(\mathbf{r}_j, \mathbf{R}_J, t) = \psi(\mathbf{r}_j, \mathbf{R}_J) \chi(\mathbf{R}_J, t) \quad (\text{D.2})$$

This is known as the Born-Oppenheimer approximation, with the electrons instantaneously updating their positions as the nuclei move. In the electronic component, the nuclei positions are parameters rather than variables, and the problem reduces to solving the time-independent Schrödinger equation,  $\hat{H}\psi = E\psi$ . However, this calculation still involves  $j$  interacting electrons, and quickly becomes intractable for large systems.

To overcome this problem, Hohenberg and Kohn showed that the ground state properties of a system are uniquely determined by the electron density  $n(r)$ , which depends on only three spatial coordinates rather than  $3j$  [345]. Minimizing an energy functional,  $F[n(r)]$ , will give the correct ground-state electron density of the system. Finally, the Kohn-Sham equations further reduce the problem from a system of interacting electrons

to a system of non-interacting electrons moving in an effective potential [344]:

$$E(n(r)) = \int V_{ext}(r) d^3r + F[n(r)] \quad (D.3)$$

The functional  $F[n(r)]$  in Equation D.3 consists of three terms:

$$F[n(r)] = T[n(r)] + E_H[n(r)] + E_{xc}[n(r)] \quad (D.4)$$

$T[n(r)]$  is the kinetic energy and  $E_H[n(r)]$  is the classical electrostatic (or Hartree) energy of the electrons. Unlike the first two terms, the exchange-correlation functional  $E_{xc}[n(r)]$  is not known exactly, and the accuracy of DFT calculations is largely dependent on this final term [346].

The simplest exchange-correlation functionals are based on the local density approximation (LDA), and assume a uniform electron gas. However, LDA functionals tend to overestimate binding energies. The generalized gradient approximation (GGA) goes a step further and also incorporates the gradient of the electron density, thus accounting for inhomogeneity in the density. As a result, GGA functionals tend to better predict properties such as total energy. Further improvements in predicting properties can be made by combining GGA functionals with a non-local Hartree-Fock exchange [347]. These are known as hybrid functionals, and often give much better descriptions of properties such as bond lengths than standard GGA approaches, although this comes at the expense of significantly longer computation times.

Lastly, density functionals are typically limited in their ability to correctly account for long-range van der Waals interactions [346]. This effect is particularly relevant in surface studies, where adsorption of atoms and molecules is often dominated by non-bonding interactions. A straightforward remedy is to apply a semi-empirical dispersion correction to the functionals derived from the electron density and reference values [348].

## D.2 Calculation details

All of the *ab initio* calculations in this work used the plane-wave density functional theory code CASTEP [349], with the Perdew-Burke-Ernzerhof (PBE) GGA functional unless otherwise stated [350]. The Tkatchenko-Scheffler semi-empirical dispersion correction scheme was used to account for van der Waals interactions [348], and all calculations were spin-polarized, starting from a uniformly distributed initial magnetic moment as implemented in CASTEP. The anionic and cationic clusters in Chapter 6 were simulated using the charge parameter implemented in CASTEP.



The isolated cluster searches with AIRSS in Chapter 6 used a supercell measuring  $14 \times 14 \times 14 \text{ \AA}$ , which was determined using convergence tests to minimize any interactions between the periodic supercell images. For the calculations involving graphene, the interlayer separation was  $14 \text{ \AA}$ , again chosen to minimize interactions between supercell images. For initial structure searches using the AIRSS and USPEX methods, the Brillouin zone was sampled at the  $\Gamma$ -point only, while final refinements of the graphene supercells used a Monkhorst-Pack sampling scheme [351], with a  $3 \times 3 \times 1$  grid following the method in Ref. 229.

A plane-wave cut-off-energy of 400 eV was used for the valence electrons in the structure searches, while the final refinements and energy calculations used a cut-off energy of 550 eV. An exception is the hybrid functional analysis of Cu clusters (Fig. 6.9), which required the use of norm-conserving rather than ultrasoft pseudopotentials, and so a larger cut-off energy of 900 eV was employed. Pseudopotentials are designed to approximate the complex interactions caused by the core electrons, under the assumption that only the valence electrons are chemically active.

Geometry optimizations were carried out using either a two-point steepest descent method (TPSD) or the quasi-Newton BFGS method as implemented in CASTEP. In all cases the systems were relaxed until the maximum force on all atoms was less than  $0.01 \text{ eV \AA}^{-1}$  and the maximum atom displacement between optimization steps was less than  $0.001 \text{ \AA}$ .

The adsorption energies of the copper adatom on graphene oxide (Chapter 5) were calculated according to the following expression:

$$E_{\text{ads}} = E_{\text{Cu/GO}} - (E_{\text{Cu}} + n_{\text{C}}E_{\text{C}} + n_{\text{O}}E_{\text{O}}) \quad (\text{D.5})$$

where  $E_{\text{Cu/GO}}$  is the energy of the optimized copper-graphene (oxide) system,  $E_{\text{Cu}}$  is the energy of an isolated copper atom,  $n_{\text{C}}$  is the number of carbon atoms in the graphene supercell, and  $E_{\text{C}}$  is the energy of a carbon atom in graphene. The number of oxygen atoms is  $n_{\text{O}}$ , and  $E_{\text{O}}$  is calculated from the energy of an isolated oxygen molecule.



# References

- [1] M. von Ardenne, “Intensitätsfragen und Auflösungsvermögen des Elektronenmikroskops [Questions concerning the intensity and resolving power of electron microscopes],” *Z. Phys.*, vol. 112, no. 11-12, pp. 744–752, 1939.
- [2] A. V. Crewe, “Scanning electron microscopes: Is high resolution possible?” *Science*, vol. 154, no. 3750, pp. 729–738, 1966.
- [3] A. V. Crewe, “A High-Resolution Scanning Transmission Electron Microscope,” *J. Appl. Phys.*, vol. 39, no. 13, p. 5861, 1968.
- [4] A. V. Crewe, J. Wall, and J. Langmore, “Visibility of Single Atoms,” *Science*, vol. 168, no. 3937, pp. 1338–1340, 1970.
- [5] S. J. Pennycook, “Seeing the atoms more clearly: STEM imaging from the Crewe era to today,” *Ultramicroscopy*, vol. 123, pp. 28–37, 2012.
- [6] B. E. Bammes *et al.*, “Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency,” *J. Struct. Biol.*, vol. 177, no. 3, pp. 589–601, 2012.
- [7] G. Zhu, G. Radtke, and G. A. Botton, “Bonding and structure of a reconstructed (001) surface of SrTiO<sub>3</sub> from TEM,” *Nature*, vol. 490, no. 7420, pp. 384–387, 2012.
- [8] I. MacLaren and Q. M. Ramasse, “Aberration-corrected scanning transmission electron microscopy for atomic-resolution studies of functional oxides,” *Int. Mater. Rev.*, vol. 59, no. 3, pp. 115–131, 2014.
- [9] E. Meijering *et al.*, “Imagining the future of bioimage analysis,” *Nat. Biotechnol.*, vol. 34, no. 12, pp. 1250–1255, 2016.
- [10] A. Zewail and J. M. Thomas, *4D Electron Microscopy: Imaging in Space and Time*, 1st ed., London: Imperial College Press, 2010.
- [11] T. LaGrange *et al.*, “Single-shot dynamic transmission electron microscopy,” *Appl. Phys. Lett.*, vol. 89, no. 4, p. 044105, 2006.
- [12] N. Browning *et al.*, “Recent developments in dynamic transmission electron microscopy,” *Curr. Opin. Solid State Mater. Sci.*, vol. 16, no. 1, pp. 23–30, 2012.
- [13] M. Isaacson *et al.*, “The study of the adsorption and diffusion of heavy atoms on light element substrates by means of the atomic resolution STEM,” *Ultramicroscopy*, vol. 1, no. 3-4, pp. 359–376, 1976.
- [14] M. Isaacson *et al.*, “Direct observations of atomic diffusion by scanning transmission electron microscopy,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 74, no. 5, pp. 1802–1806, 1977.

- [15] M. Utlaut, "Direct observation of the behavior of heavy single atoms on amorphous carbon substrates," *Phys. Rev. B*, vol. 22, no. 10, pp. 4650–4660, 1980.
- [16] D. E. Wolf and J. Villain, "Growth with Surface Diffusion," *Europhys. Lett.*, vol. 13, no. 5, pp. 389–394, 1990.
- [17] M. Copel *et al.*, "Surfactants in epitaxial growth," *Phys. Rev. Lett.*, vol. 63, no. 6, pp. 632–635, 1989.
- [18] T. Zambelli *et al.*, "Identification of the "Active Sites" of a Surface-Catalyzed Reaction," *Science*, vol. 273, no. 5282, pp. 1688–1690, 1996.
- [19] R. Metzler *et al.*, "Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking," *Phys. Chem. Chem. Phys.*, vol. 16, no. 44, pp. 24 128–24 164, 2014.
- [20] T. Schmidt *et al.*, "Imaging of single molecule diffusion." *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, no. 7, pp. 2926–2929, 1996.
- [21] S. M. A. Tabei *et al.*, "Intracellular transport of insulin granules is a subordinated random walk," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 110, no. 13, pp. 4911–4916, 2013.
- [22] J. M. Thomas, *Design and Applications of Single-Site Heterogeneous Catalysts: Contributions to Green Chemistry, Clean Technology and Sustainability*, 1st ed., London: Imperial College Press, 2012.
- [23] J. M. Thomas, "Catalysis: Tens of thousands of atoms replaced by one," *Nature*, vol. 525, no. 7569, pp. 325–326, 2015.
- [24] G. Kyriakou *et al.*, "Isolated Metal Atom Geometries as a Strategy for Selective Heterogeneous Hydrogenations," *Science*, vol. 335, no. 6073, pp. 1209–1212, 2012.
- [25] X.-F. Yang *et al.*, "Single-Atom Catalysts: A New Frontier in Heterogeneous Catalysis," *Acc. Chem. Res.*, vol. 46, no. 8, pp. 1740–1748, 2013.
- [26] M. Flytzani-Stephanopoulos, "Gold Atoms Stabilized on Various Supports Catalyze the Water–Gas Shift Reaction," *Acc. Chem. Res.*, vol. 47, no. 3, pp. 783–792, 2014.
- [27] E. C. Tyo and S. Vajda, "Catalysis by clusters with precise numbers of atoms," *Nat. Nanotech.*, vol. 10, no. 7, pp. 577–588, 2015.
- [28] S. Vajda *et al.*, "Subnanometre platinum clusters as highly active and selective catalysts for the oxidative dehydrogenation of propane," *Nat. Mater.*, vol. 8, no. 3, pp. 213–216, 2009.
- [29] A. C. Ferrari *et al.*, "Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems," *Nanoscale*, vol. 7, no. 11, pp. 4598–4810, 2015.
- [30] A. Thomas *et al.*, "Graphitic carbon nitride materials: variation of structure and morphology and their use as metal-free catalysts," *J. Mater. Chem.*, vol. 18, no. 41, p. 4893, 2008.
- [31] G. Vilé *et al.*, "A Stable Single-Site Palladium Catalyst for Hydrogenations," *Angew. Chem. Int. Ed. Engl.*, vol. 54, no. 38, pp. 11 265–11 269, 2015.

- [32] G. Gao *et al.*, "Single Atom (Pd/Pt) Supported on Graphitic Carbon Nitride as an Efficient Photocatalyst for Visible-Light Reduction of Carbon Dioxide," *J. Am. Chem. Soc.*, vol. 138, no. 19, pp. 6292–6297, 2016.
- [33] P. L. Gai *et al.*, "Visualisation of single atom dynamics and their role in nanocatalysts under controlled reaction environments," *Chem. Phys. Lett.*, vol. 592, pp. 355–359, 2014.
- [34] E. D. Boyes and P. L. Gai, "Visualising reacting single atoms under controlled conditions: Advances in atomic resolution in situ Environmental (Scanning) Transmission Electron Microscopy (E(S)TEM)," *C. R. Phys.*, vol. 15, no. 2–3, pp. 200–213, 2014.
- [35] S. Helveg *et al.*, "Observing gas-catalyst dynamics at atomic resolution and single-atom sensitivity," *Micron*, vol. 68, pp. 176–185, 2015.
- [36] E. D. Boyes and P. L. Gai, "Visualizing reacting single atoms in chemical reactions: Advancing the frontiers of materials research," *MRS Bull.*, vol. 40, no. 07, pp. 600–609, 2015.
- [37] C. Liu *et al.*, "Carbon Dioxide Conversion to Methanol over Size-Selected Cu<sub>4</sub> Clusters at Low Pressures," *J. Am. Chem. Soc.*, vol. 137, no. 27, pp. 8676–8679, 2015.
- [38] S. J. Pennycook, "Scanning transmission electron microscopy: Seeing the atoms more clearly," *MRS Bull.*, vol. 37, no. 10, pp. 943–951, 2012.
- [39] P. A. Midgley and J. M. Thomas, "Multi-Dimensional Electron Microscopy," *Angew. Chem. Int. Ed. Engl.*, vol. 53, no. 33, pp. 8614–8617, 2014.
- [40] J. M. Thomas *et al.*, "The rapidly changing face of electron microscopy," *Chem. Phys. Lett.*, vol. 631–632, pp. 103–113, 2015.
- [41] D. Williams and C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, 2nd ed., New York: Springer, 2009.
- [42] P. D. Nellist, "The principles of STEM imaging," in *Scanning Transmission Electron Microscopy*, S. J. Pennycook and P. D. Nellist, Eds., New York: Springer, 2011, pp. 91–115.
- [43] J. M. Cowley, "Image Contrast in a Transmission Scanning Electron Microscope," *Appl. Phys. Lett.*, vol. 15, no. 2, pp. 58–59, 1969.
- [44] L. Reimer and H. Kohl, *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*, 5th ed., Berlin: Springer, 2008.
- [45] J. M. Thomas and P. A. Midgley, "The modern electron microscope: A cornucopia of chemico-physical insights," *Chem. Phys.*, vol. 385, no. 1–3, pp. 1–10, 2011.
- [46] M. Haider *et al.*, "Electron microscopy image enhanced," *Nature*, vol. 392, no. 6678, pp. 768–769, 1998.
- [47] P. E. Batson, N. Dellby, and O. L. Krivanek, "Sub-ångstrom resolution using aberration corrected electron optics," *Nature*, vol. 418, no. 6898, pp. 617–620, 2002.
- [48] O. Krivanek *et al.*, "Towards sub-0.5Å electron beams," *Ultramicroscopy*, vol. 96, no. 3–4, pp. 229–237, 2003.

- [49] E. S. Reich, "Imaging hits noise barrier," *Nature*, vol. 499, no. 7457, pp. 135–136, 2013.
- [50] O. L. Krivanek *et al.*, "Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy," *Nature*, vol. 464, no. 7288, pp. 571–574, 2010.
- [51] W. Zhou *et al.*, "Single atom microscopy," *Microsc. Microanal.*, vol. 18, no. 06, pp. 1342–1354, 2012.
- [52] Q. M. Ramasse *et al.*, "Probing the Bonding and Electronic Structure of Single Atom Dopants in Graphene with Electron Energy Loss Spectroscopy," *Nano Lett.*, vol. 13, no. 10, pp. 4989–4995, 2013.
- [53] S. Findlay *et al.*, "Dynamics of annular bright field imaging in scanning transmission electron microscopy," *Ultramicroscopy*, vol. 110, no. 7, pp. 903–923, 2010.
- [54] S. D. Findlay *et al.*, "Direct Imaging of Hydrogen within a Crystalline Environment," *Appl. Phys. Express*, vol. 3, p. 116603, 2010.
- [55] R. Ishikawa *et al.*, "Direct imaging of hydrogen-atom columns in a crystal by annular bright-field electron microscopy," *Nat. Mater.*, vol. 10, no. 4, pp. 278–281, 2011.
- [56] P. D. Nellist and S. J. Pennycook, "Incoherent imaging using dynamically scattered coherent electrons," *Ultramicroscopy*, vol. 78, no. 1999, pp. 111–124, 1999.
- [57] M. D. Shannon, "Applications of Aberration-Corrected Scanning Transmission Electron Microscopy," in *Aberration-Corrected Analytical Transmission Electron Microscopy*, R. Brydson, Ed., Chichester: John Wiley & Sons, 2011, pp. 211–240.
- [58] J. M. LeBeau *et al.*, "Quantitative Atomic Resolution Scanning Transmission Electron Microscopy," *Phys. Rev. Lett.*, vol. 100, p. 206101, 2008.
- [59] S. Van Aert *et al.*, "Atomic Resolution Mapping Using Quantitative High-angle Annular Dark Field Scanning Transmission Electron Microscopy," *Microsc. Microanal.*, vol. 15, no. S2, pp. 464–465, 2009.
- [60] A. R. Lupini *et al.*, "Scanning Transmission Electron Microscopy," in *Nanocharacterisation*, A. I. Kirkland and S. J. Haigh, Eds., Cambridge: Royal Society of Chemistry, 2015, pp. 30–79.
- [61] P. B. Hirsch *et al.*, *Electron Microscopy of Thin Crystals*, 1st ed., London: Butterworths, 1965.
- [62] J. M. Cowley and A. F. Moodie, "The scattering of electrons by atoms and crystals. I. A new theoretical approach," *Acta Crystallogr.*, vol. 10, no. 10, pp. 609–619, 1957.
- [63] J. Barthel, "Dr. Probe – High-resolution (S)TEM image simulation software," 2016. [Online]. Available: <http://www.er-c.org/barthel/drprobe/>. Accessed: 01-09-2016.
- [64] C. Dwyer, R. Erni, and J. Etheridge, "Measurement of effective source distribution and its importance for quantitative interpretation of STEM images," *Ultramicroscopy*, vol. 110, no. 8, pp. 952–957, 2010.

- [65] J. Verbeeck, A. Béch , and W. V. den Broek, "A holographic method to measure the source size broadening in STEM," *Ultramicroscopy*, vol. 120, pp. 35–40, 2012.
- [66] M. Watanabe, "X-ray Energy Dispersive Spectrometry in Scanning Transmission Electron Microscopy," in *Scanning Transmission Electron Microscopy*, S. J. Pennycook and P. D. Nellist, Eds., New York: Springer, 2011, pp. 291–351.
- [67] R. Brydson and N. Hondrow, "Electron Energy Loss Spectrometry and Energy Dispersive X-ray Analysis," in *Aberration-Corrected Analytical Transmission Electron Microscopy*, R. Brydson, Ed., Chichester: John Wiley & Sons, 2011, pp. 164–210.
- [68] R. Leary *et al.*, "Compressed sensing electron tomography," *Ultramicroscopy*, vol. 131, pp. 70–91, 2013.
- [69] Z. Saghi *et al.*, "Three-Dimensional Morphology of Iron Oxide Nanoparticles with Reactive Concave Surfaces. A Compressed Sensing-Electron Tomography (CS-ET) Approach," *Nano Lett.*, vol. 11, no. 11, pp. 4666–4673, 2011.
- [70] S. V. Aert *et al.*, "Three-dimensional atomic imaging of crystalline nanoparticles," *Nature*, vol. 470, no. 7334, pp. 374–377, 2011.
- [71] L. Jones *et al.*, "Rapid Estimation of Catalyst Nanoparticle Morphology and Atomic-Coordination by High-Resolution Z-Contrast Electron Microscopy," *Nano Lett.*, vol. 14, no. 11, pp. 6336–6341, 2014.
- [72] M. Yu *et al.*, "Integrated Computational and Experimental Structure Refinement for Nanoparticles," *ACS Nano*, vol. 10, no. 4, pp. 4031–4038, 2016.
- [73] A. Doerr, "Single-particle cryo-electron microscopy," *Nature Methods*, vol. 13, no. 1, pp. 23–23, 2015.
- [74] Y. Cheng *et al.*, "A Primer to Single-Particle Cryo-Electron Microscopy," *Cell*, vol. 161, no. 3, pp. 438–449, 2015.
- [75] J. Frank, "Single-particle reconstruction of biological macromolecules in electron microscopy – 30 years," *Q. Rev. Biophys.*, vol. 42, no. 03, pp. 139–158, 2009.
- [76] X. Li *et al.*, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM," *Nat. Methods*, vol. 10, no. 6, pp. 584–590, 2013.
- [77] H. Yang *et al.*, "Imaging screw dislocations at atomic resolution by aberration-corrected electron optical sectioning," *Nat. Commun.*, vol. 6, p. 7266, 2015.
- [78] F. M. Ross, "In Situ Transmission Electron Microscopy," in *Science of Microscopy*, P. W. Hawkes and J. C. H. Spence, Eds., New York: Springer, 2007, pp. 445–534.
- [79] N. de Jonge and F. M. Ross, "Electron microscopy of specimens in liquid," *Nat. Nanotech.*, vol. 6, no. 11, pp. 695–704, 2011.
- [80] E. W. M ller and K. Bahadur, "Field Ionization of Gases at a Metal Surface and the Resolution of the Field Ion Microscope," *Phys. Rev.*, vol. 102, no. 3, pp. 624–631, 1956.
- [81] B. Gault *et al.*, *Atom Probe Microscopy*, 1st ed., New York: Springer, 2012.

- [82] T. Iwata, "Field ion microscope image of platinum," 2006. [Online]. Available: <http://en.wikipedia.org/wiki/File:FIM-platinum.jpg>. Accessed: 18-01-2017. Licensed under CC BY 2.0.
- [83] G. Antczak and G. Ehrlich, *Surface Diffusion: Metals, Metal Atoms, and Clusters*, 1st ed., Cambridge: Cambridge University Press, 2010.
- [84] G. Binnig *et al.*, "Surface studies by scanning tunneling microscopy," *Phys. Rev. Lett.*, vol. 49, no. 1, pp. 57–61, 1982.
- [85] C. Chen, *Introduction to Scanning Tunneling Microscopy*, 2nd ed., Oxford: Oxford University Press, 2008.
- [86] G. Binnig, C. F. Quate, and C. Gerber, "Atomic Force Microscope," *Phys. Rev. Lett.*, vol. 56, no. 9, pp. 930–933, 1986.
- [87] L. Gross *et al.*, "The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy," *Science*, vol. 325, no. 5944, pp. 1110–1114, 2009.
- [88] L. Gross, "Recent advances in submolecular resolution with scanning probe microscopy," *Nat. Chem.*, vol. 3, no. 4, pp. 273–278, 2011.
- [89] M. Bowker and P. R. Davies, "Preface," in *Scanning Tunneling Microscopy in Surface Science, Nanoscience and Catalysis*, M. Bowker and P. R. Davies, Eds., Weinheim: Wiley-VCH, 2010, pp. ix–xi.
- [90] F. Ercolessi, E. Tosatti, and M. Parrinello, "Au (100) Surface Reconstruction," *Phys. Rev. Lett.*, vol. 57, no. 6, pp. 719–722, 1986.
- [91] E. Rossen, "Image of surface reconstruction on a clean Gold (Au(100)) surface," 2006. [Online]. Available: [http://en.wikipedia.org/wiki/File:Atomic\\_resolution\\_Au100.JPG](http://en.wikipedia.org/wiki/File:Atomic_resolution_Au100.JPG). Accessed: 18-01-2017. Public domain image.
- [92] M. J. Rost *et al.*, "Scanning probe microscopes go video rate and beyond," *Rev. Sci. Instrum.*, vol. 76, no. 5, p. 053710, 2005.
- [93] B. S. Swartzentruber, "Direct Measurement of Surface Diffusion Using Atom-Tracking Scanning Tunneling Microscopy," *Phys. Rev. Lett.*, vol. 76, no. 3, pp. 459–462, 1996.
- [94] G. Schitter and M. J. Rost, "Scanning probe microscopy at video-rate," *Mater. Today*, vol. 11, pp. 40–48, 2008.
- [95] C. Becker and K. Wandelt, "The Template Route to Nanostructured Model Catalysts," in *Scanning Tunneling Microscopy in Surface Science, Nanoscience and Catalysis*, M. Bowker and P. R. Davies, Eds., Weinheim: Wiley-VCH, 2010, pp. 29–54.
- [96] F. Yang and D. W. Goodman, "In Situ STM Studies of Model Catalysts," in *Scanning Tunneling Microscopy in Surface Science, Nanoscience and Catalysis*, M. Bowker and P. R. Davies, Eds., Weinheim: Wiley-VCH, 2010, pp. 55–96.
- [97] M. R. Castell, "Scanning Tunnelling Microscopy of Surfaces and Nanostructures," in *Nanocharacterisation*, A. I. Kirkland and S. J. Haigh, Eds., Cambridge: Royal Society of Chemistry, 2015, pp. 80–107.



- [98] N. Shibata *et al.*, “Application of STEM to Nanomaterials and Biological Specimens,” in *Scanning transmission electron microscopy of nanomaterials: basics of imaging and analysis*, N. Tanaka, Ed., London: Imperial College Press, 2014, pp. 109–178.
- [99] J. C. Meyer *et al.*, “Accurate Measurement of Electron Beam Induced Displacement Cross Sections for Single-Layer Graphene,” *Phys. Rev. Lett.*, vol. 108, p. 196102, 2012.
- [100] J. A. Venables, *Introduction to Surface and Thin Film Processes*, 1st ed., New York: Cambridge University Press, 2000.
- [101] R. Egerton, “Beam-Induced Motion of Adatoms in the Transmission Electron Microscope,” *Microsc. Microanal.*, vol. 19, no. 02, pp. 479–486, 2013.
- [102] O. L. Krivanek *et al.*, “Gentle STEM: ADF imaging and EELS at low primary energies,” *Ultramicroscopy*, vol. 110, no. 8, pp. 935–945, 2010.
- [103] C. Kisielowski *et al.*, “Detecting structural variances of  $\text{Co}_3\text{O}_4$  catalysts by controlling beam-induced sample alterations in the vacuum of a transmission electron microscope,” *Adv. Struct. Chem. Imaging*, vol. 2, no. 1, pp. 2–13, 2016.
- [104] V. E. Cosslett, “Radiation damage in the high resolution electron microscopy of biological materials: a review,” *J. Microsc.*, vol. 113, no. 2, pp. 113–129, 1978.
- [105] M. M. Treacy, “Z Dependence of Electron Scattering by Single Atoms into Annular Dark-Field Detectors,” *Microsc. Microanal.*, vol. 17, no. 06, pp. 847–858, 2011.
- [106] J. Wall *et al.*, “Scanning Transmission Electron Microscopy at High Resolution,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 71, no. 1, pp. 1–5, 1974.
- [107] P. E. Batson, “Motion of Gold Atoms on Carbon in the Aberration-Corrected STEM,” *Microsc. Microanal.*, vol. 14, no. 01, pp. 89–97, 2008.
- [108] T. Susi *et al.*, “Silicon–Carbon Bond Inversions Driven by 60-keV Electrons in Graphene,” *Phys. Rev. Lett.*, vol. 113, p. 115501, 2014.
- [109] J. Lee *et al.*, “Direct visualization of reversible dynamics in a  $\text{Si}_6$  cluster embedded in a graphene pore,” *Nat. Commun.*, vol. 4, p. 1650, 2013.
- [110] R. Ishikawa *et al.*, “Direct Observation of Dopant Atom Diffusion in a Bulk Semiconductor Crystal Enhanced by a Large Size Mismatch,” *Phys. Rev. Lett.*, vol. 113, p. 1555051, 2014.
- [111] T. J. Pennycook *et al.*, “Atomic scale dynamics of a solid state chemical reaction directly determined by annular dark-field electron microscopy,” *Sci. Rep.*, vol. 4, p. 7555, 2014.
- [112] N. Goldman and N. D. Browning, “Gold Cluster Diffusion Kinetics on Stoichiometric and Reduced Surfaces of Rutile  $\text{TiO}_2(110)$ ,” *J. Phys. Chem. C*, vol. 115, no. 23, pp. 11 611–11 617, 2011.
- [113] R. N. Clough, G. Moldovan, and A. I. Kirkland, “Direct Detectors for Electron Microscopy,” *J. Phys. Conf. Ser.*, vol. 522, p. 012046, 2014.
- [114] A. Zewail, “Four-Dimensional Electron Microscopy,” *Science*, vol. 328, no. 5975, pp. 187–193, 2010.

- [115] A. Zewail, "Space-Time Visualization: 4D Electron Microscopy Development at Caltech," in *4D Visualization of Matter: Recent Collected Works of Ahmed H Zewail, Nobel Laureate.*, London: Imperial College Press, 2014, pp. 1–3.
- [116] X. Fu *et al.*, "Imaging rotational dynamics of nanoparticles in liquid by 4D electron microscopy," *Science*, vol. 355, no. 6324, pp. 494–498, 2017.
- [117] T. Furnival, R. K. Leary, and P. A. Midgley, "Denoising time-resolved microscopy image sequences with singular value thresholding," *Ultramicroscopy*, vol. 178, pp. 112–124, 2017.
- [118] X. Yu, D. Hu, and J. Xu, *Blind Source Separation: Theory and Applications*, 1st ed., Singapore: John Wiley & Sons, 2014.
- [119] A. Cichocki *et al.*, "Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.
- [120] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., New York: Springer, 2002.
- [121] M. Watanabe and D. Williams, "Improvements of Elemental Mapping via X-ray Spectrum Imaging Combined with Principal Component Analysis and Zero-Peak Deconvolution," *Microsc. Microanal.*, vol. 10, pp. 1040–1041, 2004.
- [122] O. Nicoletti *et al.*, "Three-dimensional imaging of localized surface plasmon resonances of metal nanoparticles," *Nature*, vol. 502, no. 7469, pp. 80–84, 2013.
- [123] S. M. Collins *et al.*, "Eigenmode tomography of surface charge oscillations of plasmonic nanoparticles by electron energy loss spectroscopy," *ACS Photonics*, vol. 2, no. 11, pp. 1628–1635, 2015.
- [124] A. S. Eggeman, R. Krakow, and P. A. Midgley, "Scanning precession electron tomography for three-dimensional nanoscale orientation imaging and crystallographic analysis," *Nat Commun.*, vol. 6, p. 7267, 2015.
- [125] A. S. Eggeman *et al.*, "Decomposing electron diffraction signals in multi-component microstructures," *Microsc. Microanal.*, vol. 21, pp. 1241–1242, 2015.
- [126] F. de la Peña *et al.*, "Mapping titanium and tin oxide phases using EELS: An application of independent component analysis," *Ultramicroscopy*, vol. 111, no. 2, pp. 169–176, 2011.
- [127] D. Rossouw *et al.*, "Multicomponent Signal Unmixing from Nanoheterostructures: Overcoming the Traditional Challenges of Nanoscale X-ray Analysis via Machine Learning," *Nano Lett.*, vol. 15, no. 4, pp. 2716–2720, 2015.
- [128] D. Rossouw *et al.*, "Blind source separation aided characterization of the  $\gamma'$  strengthening phase in an advanced nickel-based superalloy by spectroscopic 4D electron microscopy," *Acta Mater.*, vol. 107, pp. 229–238, 2016.
- [129] M. R. Keenan and P. G. Kotula, "Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images," *Surf. Interface Anal.*, vol. 36, no. 3, pp. 203–212, 2004.
- [130] S. Lichtert and J. Verbeeck, "Statistical consequences of applying a PCA noise filter on EELS spectrum images," *Ultramicroscopy*, vol. 125, pp. 35–42, 2013.
- [131] J. Salmon *et al.*, "Poisson noise reduction with non-local PCA," *J. Math. Imaging Vis.*, vol. 48, no. 2, pp. 279–294, 2013.

- [132] D. A. Ross *et al.*, “Incremental Learning for Robust Visual Tracking,” *IJCV*, vol. 77, no. 1-3, pp. 125–141, 2007.
- [133] G. Zhou *et al.*, “Efficient Nonnegative Tucker Decompositions: Algorithms and Uniqueness,” *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4990–5003, 2015.
- [134] G. Zhou *et al.*, “Linked Component Analysis From Matrices to High-Order Tensors: Applications to Biomedical Data,” *Proc. IEEE*, vol. 104, no. 2, pp. 310–331, 2016.
- [135] E. J. Candès and B. Recht, “Exact Matrix Completion via Convex Optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [136] D. L. Donoho and I. M. Johnstone, “Adapting to Unknown Smoothness via Wavelet Shrinkage,” *JASA*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [137] J.-F. Cai, E. J. Candès, and Z. Shen, “A Singular Value Thresholding Algorithm for Matrix Completion,” *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [138] S. Gu *et al.*, “Weighted Nuclear Norm Minimization with Application to Image Denoising,” in *Proc. CVPR IEEE*, IEEE, 2014, pp. 2862–2869.
- [139] Q. Xie *et al.*, “On the Optimal Solution of Weighted Nuclear Norm Minimization,” arXiv preprint, arXiv:1405.6012, 2014.
- [140] H. Ji *et al.*, “Robust video denoising using low rank matrix completion,” in *Proc. CVPR IEEE*, IEEE, 2010, pp. 1791–1798.
- [141] E. Candès, C. Sing-Long, and J. Trzasko, “Unbiased Risk Estimates for Singular Value Thresholding and Spectral Estimators,” *IEEE Trans. Sig. Process.*, vol. 61, no. 19, pp. 4643–4657, 2013.
- [142] Y. Nie and K.-K. Ma, “Adaptive rood pattern search for fast block-matching motion estimation,” *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1442–1449, 2002.
- [143] F. Luisier, T. Blu, and M. Unser, “Image Denoising in Mixed Poisson-Gaussian Noise,” *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, 2011.
- [144] Y. Le Montagner, E. Angelini, and J.-C. Olivo-Marin, “An Unbiased Risk Estimator for Image Denoising in the Presence of Mixed Poisson-Gaussian Noise,” *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1255–1268, 2014.
- [145] D. Gerard and P. Hoff, “Adaptive Higher-order Spectral Estimators,” arXiv preprint, arXiv:1505.02114, 2015.
- [146] Q. Zhao *et al.*, “Bayesian robust tensor factorization for incomplete multiway data,” *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 4, pp. 736–748, 2016.
- [147] M. Makitalo and A. Foi, “Noise Parameter Mismatch in Variance Stabilization, With an Application to Poisson-Gaussian Noise Estimation,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5348–5359, 2014.
- [148] A. Jezierska *et al.*, “An EM Approach for Time-Variant Poisson-Gaussian Model Parameter Estimation,” *IEEE Trans. Sig. Process.*, vol. 62, no. 1, pp. 17–30, 2014.
- [149] S. Ramani, C. Vonesch, and M. Unser, “Deconvolution of 3D fluorescence micrographs with automatic risk minimization,” in *ISBI 2008*, IEEE, 2008, pp. 732–735.

- [150] J. Boulanger *et al.*, “Patch-Based Nonlocal Functional for Denoising Fluorescence Microscopy Image Sequences,” *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 442–454, 2010.
- [151] N. Chenouard *et al.*, “Objective comparison of particle tracking methods,” *Nat. Methods*, vol. 11, no. 3, pp. 281–289, 2014.
- [152] M. Maggioni *et al.*, “Video Denoising, Deblocking, and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms,” *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3952–3966, 2012.
- [153] M. Maggioni and A. Foi, “V-BM4D software for video denoising, V1.0,” 2014. [Online]. Available: <http://www.cs.tut.fi/~foi/GCF-BM3D/>. Accessed: 07-10-2015.
- [154] M. Makitalo and A. Foi, “Optimal Inversion of the Generalized Anscombe Transformation for Poisson-Gaussian Noise,” *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91–103, 2013.
- [155] A. B. Yankovich *et al.*, “Picometre-precision analysis of scanning transmission electron microscopy images of platinum nanocatalysts,” *Nat. Commun.*, vol. 5, p. 4155, 2014.
- [156] L. Jones *et al.*, “Smart Align—a new tool for robust non-rigid registration of scanning microscope data,” *Adv. Struct. Chem. Imaging*, vol. 1, no. 1, pp. 1–16, 2015.
- [157] D. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, 1st ed., Cambridge: Cambridge University Press, 2003.
- [158] Perception and Decision Lab, University of Illinois, “Low-Rank Matrix Recovery and Completion via Convex Optimization,” 2012. [Online]. Available: <http://perception.csl.illinois.edu/matrix-rank/home.html>. Accessed: 08-08-2016.
- [159] E. J. Candès *et al.*, “Robust principal component analysis?” *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [160] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed., New York: Springer, 2010.
- [161] T. Zhou and D. Tao, “GoDec: Randomized Low-rank & Sparse Matrix Decomposition in Noisy Case,” in *ICML-11*, L. Getoor and T. Scheffer, Eds., ACM, 2011, pp. 33–40.
- [162] Z. Lin, R. Liu, and Z. Su, “Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation,” in *NIPS 2011*, J. Shawe-Taylor *et al.*, Eds., vol. 24., Curran Associates, 2011, pp. 612–620.
- [163] J. Feng, H. Xu, and S. Yan, “Online Robust PCA via Stochastic Optimization,” in *NIPS 2013*, C. Burges *et al.*, Eds., vol. 26., Curran Associates, 2013, pp. 404–412.
- [164] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [165] X. Shu, F. Porikli, and N. Ahuja, “Robust Orthonormal Subspace Learning: Efficient Recovery of Corrupted Low-Rank Matrices,” in *Proc. CVPR IEEE.*, IEEE, 2014, pp. 3874–3881.

- [166] J. He, L. Balzano, and A. Szlam, “Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video,” in *Proc. CVPR IEEE*, IEEE, 2012, pp. 1568–1575.
- [167] R. Zhao and V. Y. F. Tan, “Online Nonnegative Matrix Factorization with Outliers,” arXiv preprint, arXiv:1604.02634, 2016.
- [168] M. D. Zeiler, “ADADELTA: An Adaptive Learning Rate Method,” arXiv preprint, arXiv:1212.5701, 2012.
- [169] H.-G. Liao *et al.*, “Facet development during platinum nanocube growth,” *Science*, vol. 345, no. 6199, pp. 916–919, 2014.
- [170] D. Sirohi *et al.*, “The 3.8 Å resolution cryo-EM structure of Zika virus,” *Science*, vol. 352, no. 6284, pp. 467–470, 2016.
- [171] Z. Saghi *et al.*, “Reduced-dose and high-speed acquisition strategies for multi-dimensional electron microscopy,” *Adv. Struct. Chem. Imaging*, vol. 1, no. 1, pp. 1–7, 2015.
- [172] L. Kovarik *et al.*, “Implementing an accurate and rapid sparse sampling approach for low-dose atomic resolution STEM imaging,” *Appl. Phys. Lett.*, vol. 109, no. 16, p. 164102, 2016.
- [173] T. Bouwmans *et al.*, “Decomposition into Low-rank plus Additive Matrices for Background/Foreground Separation: A Review for a Comparative Evaluation with a Large-Scale Dataset,” arXiv preprint, arXiv:1511.01245, 2015.
- [174] B. Berkels *et al.*, “Optimized imaging using non-rigid registration,” *Ultramicroscopy*, vol. 138, pp. 46–56, 2014.
- [175] J. A. Högbom, “Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines,” *A&AS*, vol. 15, p. 417, 1974.
- [176] P. B. Stetson, “DAOPHOT - a computer program for crowded-field stellar photometry,” *PASP*, vol. 99, p. 191, 1987.
- [177] R. S. Savage and S. Oliver, “Bayesian Methods of Astronomical Source Extraction,” *Astrophys. J.*, vol. 661, no. 2, pp. 1339–1346, 2007.
- [178] S. Yamashita *et al.*, “Quantitative annular dark-field imaging of single-layer graphene-II: atomic-resolution image contrast,” *Microscopy (Tokyo)*, vol. 64, no. 6, pp. 409–418, 2015.
- [179] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., New York: Wiley, 2001.
- [180] C. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., New York: Springer, 2006.
- [181] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [182] J. L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, 1st ed., Cambridge: Cambridge University Press, 2010.
- [183] I. Daubechies, *Ten Lectures on Wavelets*, 1st ed., Philadelphia: SIAM, 1992.

- [184] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. CVPR IEEE.*, IEEE, 2001, pp. 511–518.
- [185] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.
- [186] A. D. Backer *et al.*, "Atom counting in HAADF STEM using a statistical model-based approach: Methodology, possibilities, and inherent limitations," *Ultramicroscopy*, vol. 134, pp. 23–33, 2013.
- [187] G. Martinez *et al.*, "Quantitative composition determination at the atomic level using model-based high-angle annular dark field scanning transmission electron microscopy," *Ultramicroscopy*, vol. 137, pp. 12–19, 2014.
- [188] H. E. *et al.*, "Probe integrated scattering cross sections in the analysis of atomic resolution HAADF STEM images," *Ultramicroscopy*, vol. 133, pp. 109–119, 2013.
- [189] K. van den Bos *et al.*, "Locating light and heavy atomic column positions with picometer precision using ISTEM," *Ultramicroscopy*, vol. 172, pp. 75–81, 2017.
- [190] F. Huang *et al.*, "Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms," *Nat. Methods*, vol. 10, no. 7, pp. 653–658, 2013.
- [191] D. Goldberg, "What every computer scientist should know about floating-point arithmetic," *ACM Comput. Surv.*, vol. 23, no. 1, pp. 5–48, 1991.
- [192] J. H. Pollard, *A handbook of numerical and statistical techniques with examples mainly from the life sciences*, 1st ed., Cambridge: Cambridge University Press, 1977.
- [193] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: a review of algorithms and comparison of software implementations," *J. Global Optim.*, vol. 56, no. 3, pp. 1247–1293, 2012.
- [194] A. Bos, *Parameter Estimation for Scientists and Engineers*, 1st ed., Hoboken, NJ: John Wiley & Sons, 2007.
- [195] M. Alania *et al.*, "How precise can atoms of a nanocluster be located in 3D using a tilt series of scanning transmission electron microscopy images?" *Ultramicroscopy*, 2016, in press.
- [196] H. H. Barrett, C. Dainty, and D. Lara, "Maximum-likelihood methods in wave-front sensing: stochastic models and likelihood functions," *J. Opt. Soc. Am. A*, vol. 24, no. 2, p. 391, 2007.
- [197] C. S. Smith *et al.*, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," *Nat. Methods*, vol. 7, no. 5, pp. 373–375, 2010.
- [198] A. D. Backer *et al.*, "Dose limited reliability of quantitative annular dark field scanning transmission electron microscopy for nano-particle atom-counting," *Ultramicroscopy*, vol. 151, pp. 56–61, 2015.
- [199] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control.*, vol. 19, no. 6, pp. 716–723, 1974.
- [200] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

- [201] N. Hagen, M. Kupinski, and E. L. Dereniak, "Gaussian profile estimation in one dimension," *Appl. Optics*, vol. 46, no. 22, p. 5374, 2007.
- [202] F. Huang *et al.*, "Simultaneous multiple-emitter fitting for single molecule super-resolution imaging," *Biomed. Opt. Express*, vol. 2, no. 5, p. 1377, 2011.
- [203] J. C. Crocker and D. G. Grier, "Methods of Digital Video Microscopy for Colloidal Studies," *J. Colloid Interface Sci.*, vol. 179, no. 1, pp. 298–310, 1996.
- [204] I. Sbalzarini and P. Koumoutsakos, "Feature point tracking and trajectory analysis for video imaging in cell biology," *J. Struct. Bio.*, vol. 151, no. 2, pp. 182–195, 2005.
- [205] N. Chenouard, I. Bloch, and J. Olivo-Marin, "Multiple Hypothesis Tracking for Cluttered Biological Image Sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2736–3750, 2013.
- [206] F. de Chaumont *et al.*, "Icy: an open bioimage informatics platform for extended reproducible research," *Nat. Methods*, vol. 9, no. 7, pp. 690–696, 2012.
- [207] D. Silver, "Introduction to Reinforcement Learning," 2015. [Online]. Available: [http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching\\_files/intro\\_RL.pdf](http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching_files/intro_RL.pdf). Accessed: 01-10-2016.
- [208] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.
- [209] D. Barber, *Bayesian Reasoning and Machine Learning*, 2nd ed., Cambridge: Cambridge University Press, 2012.
- [210] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Comm. ACM*, vol. 18, no. 9, pp. 509–517, 1975.
- [211] T. Ala-Nissila, R. Ferrando, and S. C. Ying, "Collective and single particle diffusion on surfaces," *Adv. Phys.*, vol. 51, no. 3, pp. 949–1078, 2002.
- [212] A. Chatterjee and D. G. Vlachos, "An overview of spatial microscopic and accelerated kinetic Monte Carlo methods," *J. Comput. Aided Mater.*, vol. 14, no. 2, pp. 253–308, 2007.
- [213] T. Furnival *et al.*, "Anomalous Diffusion of Single Metal Atoms on a Graphene Oxide Support," pp. 370–374, 2017.
- [214] G. Eda, G. Fanchini, and M. Chhowalla, "Large-area ultrathin films of reduced graphene oxide as a transparent and flexible electronic material," *Nat. Nanotech.*, vol. 3, no. 5, pp. 270–274, 2008.
- [215] K. P. Loh *et al.*, "Graphene oxide as a chemically tunable platform for optical applications," *Nat. Chem.*, vol. 2, no. 12, pp. 1015–1024, 2010.
- [216] G. Eda and M. Chhowalla, "Chemically Derived Graphene Oxide: Towards Large-Area Thin-Film Electronics and Optoelectronics," *Adv. Mater.*, vol. 22, no. 22, pp. 2392–2415, 2010.
- [217] A. Tararan *et al.*, "Revisiting Graphene Oxide Chemistry via Spatially-Resolved Electron Energy Loss Spectroscopy," *Chem. Mater.*, vol. 28, no. 11, pp. 3741–3748, 2016.

- [218] X. Chen *et al.*, "Synthesis of "Clean" and Well-Dispersive Pd Nanoparticles with Excellent Electrocatalytic Property on Graphene Oxide," *J. Am. Chem. Soc.*, vol. 133, no. 11, pp. 3693–3695, 2011.
- [219] F. Li, J. Zhao, and Z. Chen, "Fe-Anchored Graphene Oxide: A Low-Cost and Easily Accessible Catalyst for Low-Temperature CO Oxidation," *J. Phys. Chem. C*, vol. 116, no. 3, pp. 2507–2514, 2012.
- [220] C. Yin *et al.*, "Atomically precise (catalytic) particles synthesized by a novel cluster deposition instrument," *J. Chem. Phys.*, vol. 140, no. 17, p. 174201, 2014.
- [221] D. K. Kim and E. Iglesia, "Isotopic and Kinetic Assessment of the Mechanism of CH<sub>3</sub>OH-H<sub>2</sub>O Catalysis on Supported Copper Clusters," *J. Phys. Chem. C*, vol. 112, no. 44, pp. 17 235–17 243, 2008.
- [222] C. Shi *et al.*, "Copper Oxide Clusters Stabilized by Ceria for CO, C<sub>3</sub>H<sub>6</sub>, and NO Abatement," *Chinese J. Catal.*, vol. 33, no. 9-10, pp. 1455–1462, 2012.
- [223] S. Hirabayashi and M. Ichihashi, "Catalytic oxidation of CO with N<sub>2</sub>O on isolated copper cluster anions," *Phys. Chem. Chem. Phys.*, vol. 16, no. 48, pp. 26 500–26 505, 2014.
- [224] K. Erickson *et al.*, "Determination of the Local Chemical Structure of Graphene Oxide and Reduced Graphene Oxide," *Adv. Mater.*, vol. 22, no. 40, pp. 4467–4472, 2010.
- [225] C. Gomez-Navarro *et al.*, "Atomic Structure of Reduced Graphene Oxide," *Nano Lett.*, vol. 10, no. 4, pp. 1144–1148, 2010.
- [226] M.-T. Nguyen, R. Erni, and D. Passerone, "Two-dimensional nucleation and growth mechanism explaining graphene oxide structures," *Phys. Rev. B*, vol. 86, p. 115406, 2012.
- [227] S. H. Dave *et al.*, "Chemistry and Structure of Graphene Oxide via Direct Imaging," *ACS Nano*, vol. 10, no. 8, pp. 7515–7522, 2016.
- [228] M. Amft *et al.*, "Adsorption of Cu, Ag, and Au atoms on graphene including van der Waals interactions," *J. Phys. Condens. Matter*, vol. 23, no. 39, p. 395001, 2011.
- [229] T. P. Hardcastle *et al.*, "Mobile metal adatoms on single layer, bilayer, and trilayer graphene: An ab initio DFT study with van der Waals corrections correlated with electron microscopy data," *Phys. Rev. B*, vol. 87, p. 195430, 2013.
- [230] A. W. Robertson *et al.*, "Dynamics of Single Fe Atoms in Graphene Vacancies," *Nano Lett.*, vol. 13, no. 4, pp. 1468–1475, 2013.
- [231] C. J. Pickard and R. J. Needs, "High-Pressure Phases of Silane," *Phys. Rev. Lett.*, vol. 97, p. 045504, 2006.
- [232] C. J. Pickard and R. J. Needs, "Ab initio random structure searching," *J. Phys. Condens. Matter*, vol. 23, no. 5, p. 053201, 2011.
- [233] X.-B. Li *et al.*, "Adaptive cluster expansion approach for predicting the structure evolution of graphene oxide," *J. Chem. Phys.*, vol. 141, no. 22, p. 224703, 2014.



- [234] S. A. Bradley *et al.*, “Behavior of Pt Atoms on Oxide Supports During Reduction Treatments at Elevated Temperatures, Characterized by Aberration Corrected Stem Imaging,” *Catal. Lett.*, vol. 142, no. 2, pp. 176–182, 2011.
- [235] J. Zhao *et al.*, “Direct in situ observations of single Fe atom catalytic processes and anomalous diffusion at graphene edges,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 44, pp. 15 641–15 646, 2014.
- [236] R. Brown, “A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies,” *Philos. Mag. S. 2*, vol. 4, no. 21, pp. 161–173, 1828.
- [237] A. Einstein, “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen [On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat],” *Ann. Phys.*, vol. 322, no. 8, pp. 549–560, 1905.
- [238] K. Pearson, “The Problem of the Random Walk,” *Nature*, vol. 72, no. 1865, p. 294, 1905.
- [239] L. Bachelier *et al.*, *Louis Bachelier’s Theory of Speculation: the Origins of Modern Finance.*, Princeton: Princeton University Press, 2006.
- [240] E. Nelson, *Dynamical theories of Brownian motion*, 2nd ed., Princeton: Princeton University Press, 2001.
- [241] W. Paul and J. Baschnagel, *Stochastic Processes: From Physics to Finance*, 1st ed., Berlin: Springer, 1999.
- [242] M. von Smoluchowski, “Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen [On the kinetic theory of Brownian molecular movement and suspensions],” *Ann. Phys.*, vol. 326, no. 14, pp. 756–780, 1906.
- [243] J. Perrin, “Mouvement brownien et réalité moléculaire [Brownian movement and molecular reality],” *Ann. Chim. Phys. 8<sup>me</sup> Series*, 1910.
- [244] Y. He *et al.*, “Random time-scale invariant diffusion and transport coefficients,” *Phys. Rev. Lett.*, vol. 101, p. 058101, 2008.
- [245] I. Golding and E. C. Cox, “Physical nature of bacterial cytoplasm,” *Phys. Rev. Lett.*, vol. 96, p. 098102, 2006.
- [246] A. V. Weigel *et al.*, “Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 16, pp. 6438–6443, 2011.
- [247] B. Mandelbrot, *The Fractal Geometry of Nature*, 1st ed., San Francisco: W. H. Freeman and Company, 1982.
- [248] V. Zaburdaev, S. Denisov, and J. Klafter, “Lévy walks,” *Rev. Mod. Phys.*, vol. 87, no. 2, pp. 483–530, 2015.
- [249] E. W. Montroll and G. H. Weiss, “Random walks on lattices. II,” *J. Math. Phys.*, vol. 6, no. 2, p. 167, 1965.
- [250] S. Havlin and D. Ben-Avraham, “Diffusion in disordered media,” *Adv. Phys.*, vol. 51, no. 1, pp. 187–292, 2002.

- [251] B. Mandelbrot and J. W. V. Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.
- [252] J. H. P. Schulz, A. V. Chechkin, and R. Metzler, "Correlated continuous time random walks: combining scale-invariance with long-range memory for spatial and temporal dynamics," *J. Phys. A*, vol. 46, no. 47, p. 475001, 2013.
- [253] M. Magdziarz *et al.*, "Fractional Brownian Motion Versus the Continuous-Time Random Walk: A Simple Test for Subdiffusive Dynamics," *Phys. Rev. Lett.*, vol. 103, p. 180602, 2009.
- [254] A. F. Voter, "Introduction to the Kinetic Monte Carlo Method," in *Radiation Effects in Solids*, K. E. Sickafus, E. A. Kotomin, and B. P. Uberuaga, Eds., Springer, 2007, pp. 1–23.
- [255] B. Mandelbrot, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension," *Science*, vol. 156, no. 3775, pp. 636–638, 1967.
- [256] J. Gleick, *Chaos: making a new science*, 2nd ed., London: Minerva, 1997.
- [257] R. F. Voss, R. B. Laibowitz, and E. I. Alessandrini, "Fractal (Scaling) Clusters in Thin Gold Films near the Percolation Threshold," *Phys. Rev. Lett.*, vol. 49, no. 19, pp. 1441–1444, 1982.
- [258] A. Kapitulnik and G. Deutscher, "Percolation Characteristics in Discontinuous Thin Films of Pb," *Phys. Rev. Lett.*, vol. 49, no. 19, pp. 1444–1448, 1982.
- [259] C. Mattevi *et al.*, "Evolution of Electrical, Chemical, and Structural Properties of Transparent and Conducting Chemically Derived Graphene Thin Films," *Adv. Funct. Mater.*, vol. 19, no. 16, pp. 2577–2583, 2009.
- [260] Y. Meroz, I. M. Sokolov, and J. Klafter, "Subdiffusion of mixed origins: When ergodicity and nonergodicity coexist," *Phys. Rev. E*, vol. 81, no. 1, p. 010101, 2010.
- [261] S. Vajda, Personal communication, December 2016.
- [262] Z. Chen *et al.*, "Stabilization of Single Metal Atoms on Graphitic Carbon Nitride," *Adv. Func. Mater.*, vol. 27, no. 8, p. 1605785, 2017.
- [263] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. ICCV IEEE.*, IEEE, 1998, pp. 839–846.
- [264] K. J. Falconer, *Techniques in Fractal Geometry*, 1st ed., New York: Wiley, 1997.
- [265] D. ben-Avraham and S. Havlin, *Diffusion and reactions in fractals and disordered systems*, 2nd ed., Cambridge: Cambridge University Press, 2005.
- [266] S. van der Walt *et al.*, "scikit-image: image processing in Python," *PeerJ*, vol. 2, p. e453, 2014.
- [267] A. Baddeley, "Spatial Point Processes and their Applications," in *Stochastic Geometry*, W. Weil, Ed., Berlin: Springer, 2007, pp. 1–75.
- [268] B. D. Ripley, "The Second-Order Analysis of Stationary Point Processes," *J. Appl. Probab.*, vol. 13, no. 2, p. 255, 1976.
- [269] T. E. Smith, "Notebook on spatial data analysis," 2016. [Online]. Available: <http://www.seas.upenn.edu/~ese502/#notebook>. Accessed: 31-10-2016.

- [270] K. Sohlberg *et al.*, “Origin of Anomalous Pt–Pt Distances in the Pt/Alumina Catalytic System,” *ChemPhysChem*, vol. 5, no. 12, pp. 1893–1897, 2004.
- [271] D. Wolpert and W. Macready, “No free lunch theorems for optimization,” *IEEE Trans. Evolut. Comput.*, vol. 1, no. 1, pp. 67–82, 1997.
- [272] J. C. Schön, K. Doll, and M. Jansen, “Predicting solid compounds via global exploration of the energy landscape of solids on the ab initio level without recourse to experimental information,” *Phys. Status Solidi B*, vol. 247, no. 1, pp. 23–39, 2010.
- [273] D. J. Wales and J. P. K. Doye, “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms,” *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–5116, 1997.
- [274] D. J. Wales and H. A. Scheraga, “Global Optimization of Clusters, Crystals, and Biomolecules,” *Science*, vol. 285, no. 5432, pp. 1368–1372, 1999.
- [275] S. Goedecker, “Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems,” *J. Chem. Phys.*, vol. 120, no. 21, p. 9911, 2004.
- [276] C. W. Glass, A. R. Oganov, and N. Hansen, “USPEX—Evolutionary crystal structure prediction,” *Comput. Phys. Commun.*, vol. 175, no. 11–12, pp. 713–720, 2006.
- [277] S. E. Schöborn *et al.*, “The performance of minima hopping and evolutionary algorithms for cluster structure prediction,” *J. Chem. Phys.*, vol. 130, no. 14, p. 144108, 2009.
- [278] A. O. Lyakhov *et al.*, “New developments in evolutionary structure prediction algorithm USPEX,” *Comput. Phys. Commun.*, vol. 184, no. 4, pp. 1172–1182, 2013.
- [279] C. Ophus *et al.*, “Large-scale experimental and theoretical study of graphene grain boundary structures,” *Phys. Rev. B*, vol. 92, p. 205402, 2015.
- [280] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [281] D. N. Johnstone *et al.*, “Mesoscale structure and topography of chemical vapour deposited graphene,” *Manuscript in preparation*.
- [282] G. Schusteritsch and C. J. Pickard, “Predicting interface structures: From SrTiO<sub>3</sub> to graphene,” *Phys. Rev. B*, vol. 90, p. 035424, 2014.
- [283] M. Itoh *et al.*, “Comprehensive study of sodium, copper, and silver clusters over a wide range of sizes  $2 \leq n \leq 75$ ,” *J. Chem. Phys.*, vol. 131, no. 17, p. 174510, 2009.
- [284] G. H. Guvelioglu *et al.*, “Evolution of Small Copper Clusters and Dissociative Chemisorption of Hydrogen,” *Phys. Rev. Lett.*, vol. 94, p. 026103, 2005.
- [285] D. Die *et al.*, “Insights into the structural, electronic and magnetic properties of V-doped copper clusters: comparison with pure copper clusters,” *Sci. Rep.*, vol. 6, p. 31978, 2016.
- [286] S. Klacar *et al.*, “Oxidation of Small Silver Clusters: A Density Functional Theory Study,” *J. Phys. Chem. C*, vol. 114, no. 29, pp. 12 610–12 617, 2010.

- [287] F. Furche *et al.*, “The structures of small gold cluster anions as determined by a combination of ion mobility measurements and density functional calculations,” *J. Chem. Phys.*, vol. 117, no. 15, p. 6982, 2002.
- [288] M. Sakurai *et al.*, “Magic numbers in transition metal (Fe, Ti, Zr, Nb, and Ta) clusters observed by time-of-flight mass spectrometry,” *J. Chem. Phys.*, vol. 111, no. 1, pp. 235–238, 1999.
- [289] C. M. Chang and M. Y. Chou, “Alternative Low-Symmetry Structure for 13-Atom Metal Clusters,” *Phys. Rev. Lett.*, vol. 93, p. 133401, 2004.
- [290] C. Barrón *et al.*, “A genetic algorithm for Lennard-Jones atomic clusters,” *Appl. Math. Lett.*, vol. 12, no. 7, pp. 85–90, 1999.
- [291] A. van de Walle and G. Ceder, “Correcting overbinding in local-density-approximation calculations,” *Phys. Rev. B*, vol. 59, no. 23, pp. 14 992–15 001, 1999.
- [292] M. G. Medvedev *et al.*, “Density functional theory is straying from the path toward the exact functional,” *Science*, vol. 355, no. 6320, pp. 49–52, 2017.
- [293] V. E. Matulis, O. A. Ivashkevich, and V. S. Gurin, “DFT study of electronic structure and geometry of anionic copper clusters,” *Comp. Theor. Chem.*, vol. 681, no. 1-3, pp. 169–176, 2004.
- [294] G.-T. Bae, B. Dellinger, and R. W. Hall, “Density Functional Calculation of the Structure and Electronic Properties of  $\text{Cu}_n\text{O}_n$  ( $n = 1-8$ ) Clusters,” *J. Phys. Chem. A*, vol. 115, no. 11, pp. 2087–2095, 2011.
- [295] P. A. Crozier, T. Aoki, and Q. Liu, “Detection of water and its derivatives on individual nanoparticles using vibrational electron energy-loss spectroscopy,” *Ultramicroscopy*, vol. 169, pp. 30–36, 2016.
- [296] T. Furnival *et al.*, “Adatom dynamics and the surface reconstruction of Si(110) revealed using time-resolved electron microscopy,” *Manuscript submitted*.
- [297] J. Wintterlin *et al.*, “Real-time STM observations of atomic equilibrium fluctuations in an adsorbate system: O/Ru(0001),” *Surf. Sci.*, vol. 394, no. 1-3, pp. 159–169, 1997.
- [298] W. Xu *et al.*, “In-situ real-space imaging of single crystal surface reconstructions via electron microscopy,” *Appl. Phys. Lett.*, vol. 109, no. 20, p. 201601, 2016.
- [299] G. Binnig *et al.*, “ $7\times 7$  Reconstruction on Si(111) Resolved in Real Space,” *Phys. Rev. Lett.*, vol. 50, no. 2, pp. 120–123, 1983.
- [300] J. Ihm, M. L. Cohen, and D. J. Chadi, “ $(2\times 1)$  reconstructed Si(001) surface: Self-consistent calculations of dimer models,” *Phys. Rev. B*, vol. 21, no. 10, pp. 4592–4599, 1980.
- [301] T. An *et al.*, “Elemental structure in Si(110)- $(16\times 2)$  revealed by scanning tunneling microscopy,” *Phys. Rev. B*, vol. 61, no. 4, pp. 3006–3011, 2000.
- [302] A. A. Stekolnikov, J. Furthmüller, and F. Bechstedt, “Long-Range Surface Reconstruction: Si(110)- $(16\times 2)$ ,” *Phys. Rev. Lett.*, vol. 93, p. 136104, 2004.
- [303] T. Yamasaki *et al.*, “First-principles theory of Si(110)- $(16\times 2)$  surface reconstruction for unveiling origin of pentagonal scanning tunneling microscopy images,” *Appl. Phys. Express*, vol. 9, p. 035501, 2016.

- [304] H. Togashi *et al.*, "Observation of Initial Oxidation on Si(110)- $16\times 2$  surface by Scanning Tunneling Microscopy," *Jpn. J. Appl. Phys.*, vol. 46, no. 5B, pp. 3239–3243, 2007.
- [305] M. Suemitsu *et al.*, "Initial oxidation of Si(110) as studied by real-time synchrotron-radiation X-ray photoemission spectroscopy," *J. Vac. Sci. Technol.*, vol. 27, no. 1, p. 547, 2009.
- [306] T. Nagasawa, S. Shiba, and K. Sueoka, "First-principles study on initial stage of oxidation on Si(110) surface," *Phys. Status Solidi C*, vol. 8, no. 3, pp. 717–720, 2011.
- [307] K. Raghavachari, "First-Principles Quantum Chemical Investigations of Silicon Oxidation," in *Fundamental Aspects of Silicon Oxidation*, Y. Chabal, Ed., New York: Springer, 2001, pp. 127–141.
- [308] B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon," *J. Appl. Phys.*, vol. 36, no. 12, p. 3770, 1965.
- [309] S. T. Pantelides *et al.*, "Local and Global Bonding at the Si–SiO<sub>2</sub> Interface," in *Fundamental Aspects of Silicon Oxidation*, Y. Chabal, Ed., New York: Springer, 2001, pp. 193–218.
- [310] Y. Chabal, Ed., *Fundamental Aspects of Silicon Oxidation.*, New York: Springer, 2001.
- [311] K. Oura *et al.*, "Hydrogen interaction with clean and modified silicon surfaces," *Surf. Sci. Rep.*, vol. 35, no. 1-2, pp. 1–69, 1999.
- [312] G. S. Chen, C. B. Boothroyd, and C. J. Humphreys, "Electron-beam-induced damage in amorphous SiO<sub>2</sub> and the direct fabrication of silicon nanostructures," *Philos. Mag. A*, vol. 78, no. 2, pp. 491–506, 1998.
- [313] H. Tokumoto and M. Iwatsuki, "Scanning Tunneling Microscopy of Clean Silicon Surfaces at Elevated Temperatures," *Jpn. J. Appl. Phys.*, vol. 32, no. Part 1, No. 3B, pp. 1368–1378, 1993.
- [314] L. V. Arapkina *et al.*, "Phase transition on the si(001) clean surface prepared in UHV MBE chamber: a study by high-resolution STM and in situ RHEED," *Nanoscale Res. Lett.*, vol. 6, no. 1, p. 218, 2011.
- [315] H. Kato *et al.*, "Preparation of an Ultraclean and Atomically Controlled Hydrogen-Terminated Si(111)-( $1\times 1$ ) Surface Revealed by High Resolution Electron Energy Loss Spectroscopy, Atomic Force Microscopy, and Scanning Tunneling Microscopy: Aqueous NH<sub>4</sub>F Etching Process of Si(111)," *Jpn. J. Appl. Phys.*, vol. 46, no. 9A, pp. 5701–5705, 2007.
- [316] J. Kim *et al.*, "Preparation of atomically clean and flat si(100) surfaces by low-energy ion sputtering and low-temperature annealing," *Appl. Surf. Sci.*, vol. 220, no. 1-4, pp. 293–297, 2003.
- [317] A. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed., Harlow, England: Prentice Hall, 2001.
- [318] F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," *Phys. Rev. B*, vol. 31, no. 8, pp. 5262–5271, 1985.

- [319] J. Tersoff, "Empirical interatomic potential for silicon with improved elastic properties," *Phys. Rev. B*, vol. 38, no. 14, pp. 9902–9905, 1988.
- [320] M. I. Baskes, "Modified embedded-atom potentials for cubic materials and impurities," *Phys. Rev. B*, vol. 46, no. 5, pp. 2727–2742, 1992.
- [321] B. Jelinek *et al.*, "Modified embedded atom method potential for Al, Si, Mg, Cu, and Fe alloys," *Phys. Rev. B*, vol. 85, p. 245102, 2012.
- [322] A. C. T. van Duin *et al.*, "ReaxFF: A Reactive Force Field for Hydrocarbons," *J. Phys. Chem. A*, vol. 105, no. 41, pp. 9396–9409, 2001.
- [323] A. D. Kulkarni *et al.*, "Oxygen Interactions with Silica Surfaces: Coupled Cluster and Density Functional Investigation and the Development of a New ReaxFF Potential," *J. Phys. Chem. C*, vol. 117, no. 1, pp. 258–269, 2013.
- [324] L. C. Ciacchi and M. C. Payne, "First-Principles Molecular-Dynamics Study of Native Oxide Growth on Si(001)," *Phys. Rev. Lett.*, vol. 95, p. 196101, 2005.
- [325] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *J. Comput. Phys.*, vol. 117, no. 1, pp. 1–19, 1995.
- [326] H. Aktulga *et al.*, "Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques," *Parallel Comput.*, vol. 38, no. 4–5, pp. 245–259, 2012.
- [327] M. A. Pamungkas *et al.*, "Reactive molecular dynamics simulation of early stage of dry oxidation of Si (100) surface," *J. Appl. Phys.*, vol. 110, no. 5, p. 053513, 2011.
- [328] X. Sang and J. M. LeBeau, "Revolving scanning transmission electron microscopy: Correcting sample drift distortion without prior knowledge," *Ultramicroscopy*, vol. 138, pp. 28–35, 2014.
- [329] G. Brumfiel, "High-energy physics: Down the petabyte highway," *Nature*, vol. 469, no. 7330, pp. 282–283, 2011.
- [330] S. Bals *et al.*, "Atomic scale dynamics of ultrasmall germanium clusters," *Nat. Commun.*, vol. 3, p. 897, 2012.
- [331] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 878–892, 2008.
- [332] F. Dellaert *et al.*, "Structure from motion without correspondence," in *Proc. CVPR IEEE*, vol. 2, 2000, pp. 557–564.
- [333] N. P. E. Barry *et al.*, "Osmium Atoms and Os<sub>2</sub> Molecules Move Faster on Selenium-Doped Compared to Sulfur-Doped Boronic Graphenic Surfaces," *Chem. Mater.*, vol. 27, no. 14, pp. 5100–5105, 2015.
- [334] S. V. Kalinin, A. Borisevich, and S. Jesse, "Fire up the atom forge," *Nature*, vol. 539, no. 7630, pp. 485–487, 2016.
- [335] N. P. E. Barry *et al.*, "Fabrication of crystals from single metal atoms," *Nat. Commun.*, vol. 5, p. 3851, 2014.
- [336] F. E. Kalff *et al.*, "A kilobyte rewritable atomic memory," *Nat. Nanotech.*, vol. 11, no. 11, pp. 926–929, 2016.

- [337] C. Sanderson and R. Curtin, “Armadillo: a template-based C++ library for linear algebra,” *J. Open Source Software*, vol. 1, no. 2, pp. 1–2, 2016.
- [338] F. de la Peña *et al.*, “HyperSpy: HyperSpy 1.1,” 2016. [Online]. Available: <http://www.hyperspy.org>. Accessed: 14-09-2016.
- [339] M. E. J. Newman and R. M. Ziff, “Fast Monte Carlo algorithm for site or bond percolation,” *Phys. Rev. E*, vol. 64, no. 1, p. 016706, 2001.
- [340] J. T. Hwang, “Improving Upon Standard Estimators in Discrete Exponential Families with Applications to Poisson and Negative Binomial Cases,” *Ann. Statist.*, vol. 10, no. 3, pp. 857–867, 1982.
- [341] L. Bottou and O. Bousquet, “The tradeoffs of large scale learning,” in *NIPS 2008*, J. Platt *et al.*, Eds., vol. 20., Curran Associates, 2007, pp. 161–168.
- [342] K. B. Petersen and M. S. Pedersen, “The Matrix Cookbook,” 2012. [Online]. Available: [http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=3274](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274). Accessed: 11-07-2016.
- [343] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv preprint, arXiv:1609.04747, 2016.
- [344] R. Zeller, “Introduction to Density Functional Theory,” in *37<sup>th</sup> IFF Spring School 2006 Computational Condensed Matter Physics*, S. Blügel *et al.*, Eds., Jülich: Forschungszentrum Jülich, 2006, pp. 2–19.
- [345] P. Hohenberg and W. Kohn, “Inhomogeneous Electron Gas,” *Phys. Rev.*, vol. 136, no. 3B, pp. B864–B871, 1964.
- [346] J. R. B. Gomes *et al.*, “Density Functional Treatment of Interactions and Chemical Reactions at Interfaces,” in *Density Functional Theory: Principles, Applications and Analysis*, J. Morin and J. M. Pelletier, Eds., New York: Nova, 2013, pp. 1–59.
- [347] J. P. Perdew, M. Ernzerhof, and K. Burke, “Rationale for mixing exact exchange with density functional approximations,” *J. Chem. Phys.*, vol. 105, no. 22, pp. 9982–9985, 1996.
- [348] A. Tkatchenko and M. Scheffler, “Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data,” *Phys. Rev. Lett.*, vol. 102, p. 073005, 2009.
- [349] S. J. Clark *et al.*, “First principles methods using CASTEP,” *Z. Kristallogr.*, vol. 220, no. 5/6, pp. 567–570, 2005.
- [350] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple,” *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868, 1997.
- [351] H. J. Monkhorst and J. D. Pack, “Special points for Brillouin-zone integrations,” *Phys. Rev. B*, vol. 13, no. 12, pp. 5188–5192, 1976.

